Role of Data Mining in Internet of Things with Challenges and Application

[1]Abha Jain, [2]Mily Lal, [3]Mohini Avatade
[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor,
Computer Department, DYPIEMR, Pune
[1]Computer Department
[1]DYPIEMR, Pune, India

Abstract : The Internet of Things (IoT) visualizes a future in which everything on the earth digital and physical objects (e.g., smart phones, TVs, cars) can be connected together via internet. The IoT's characteristics, including a global network infrastructure with self configuring capabilities of things and device with network level heterogeneity and intelligent interface, all will make development of the diverse applications and services, a very challenging task.
Firstly, Mining information and knowledge from large databases that is generated or captured by IoT, has been recognized very useful for many applications. Secondly, Data mining is playing a critical role in making smart system to provide more convenient services and environments, Data Mining also provide too many researchers as a key research topic that is associated with the development of IOT enabling technologies, infrastructure and applications. This survey begins with a introduction of Internet of things. And a brief review of the "role of data mining in IoT", in which we discussed a Data mining model for IOT to provide a smart world. In addition challenges, applications and future trends of this field are highlighted.

IndexTerms **- distributed processing, Internet of Things (IoT) applications, data mining, knowledge discovery,**

I. INTRODUCTION

The Internet of Things (IoT) refers to future generation of Internet which will include smart environment of uniquely identifiable object and self-aware things to handle large web servers and supercomputer clusters and build a new smart world [1][32][36].

This future internet will create the new revolution in the field of information technology after the revolution of computer and Internet. The basic idea of this concept is the pervasive presence around us of a variety of things or objects which are capable to interact with each other to perform a specific task and manageable by their own in case of failure.[2][29].
.
S. Haller et al.[3] have given the following definition of IoT : "A world where physical objects are seamlessly integrated into the information network, and where the physical objects can become active participants in business process. Services are available to interact with these 'smart object' over the Internet, query their state and any information associated with them, taking into account security and privacy issues."

IOT is not only a technology but it is a concept ,which is applicable to smallest area such as home ,street light signal to a wider area such as real time application to support ,assets and providing new services. These are providing lot of business opportunities but also adding complexity to it.

The internet of thing are creating large amount of the data and it will be create more data in future also.So ,it is necessary to developed methods to effectively handle the data, analyze it and mining it.IoT data can be categorized into several types as descriptive data, environmental data ,positional data RFID data stream and sensor network data[4]. It brings the great challenges for managing , analyzing and mining different types of IOT data.

Recently, the large amount of data collected and stored around the world has been increasing at the exponential rate. IOT produced data which is heterogeneous and large scale, because it is produced by different type of IoT devices like sensors, networks, the web, social media and transactional devices. the traditional methods and technologies are inadequate to deal with data storage and analyzing the complex and large scale data [39]. There are various latestest technology which are used to store, process and analyze this vast amount of heterogeneous data.[39] For example Cloud-based Technologies, machine learning algorithms, Artifical Intelligence, Big Data are latestest technologies which have been voluminous and complex data. So this is required to introduce intelligent ways of data analysis

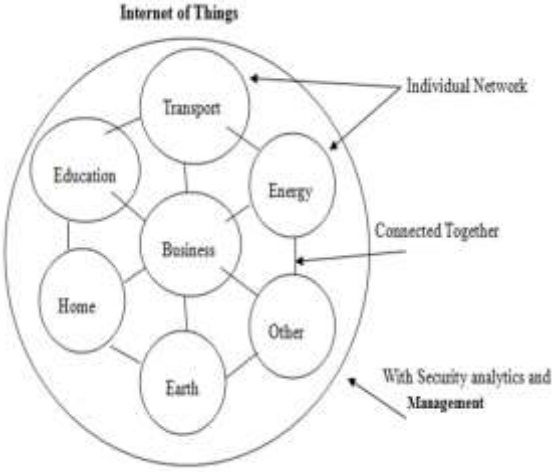

Fig 1 : IOT viewed as a network of networks

that is Data Mining methods [39]Data Mining is extraction of the information or knowlede from the hidden data set.The goal of data mining is to make massive data effective, and easily understandable for data scientists and business intelligence professionals. Thus, the simulation and solving of real-world time- and resource-consuming data mining problems may be realized.

This paper gives a overview of Internet of things and the role of data mining in IOT in present time alsoin next generation internet environment. section ii cover the literature survey. In III section , the focus on development on effective methods of data processing for managing, analyzing and mining data. Iv section cover the challenges of data mining toward IOT technologies for IoT . Section V and VI Present the Next generation IOT and Data Mining also application of IOT .Conclusions and future trends are drawn in last Section VII

2. RELATED WORK

Internet of thing has become part of your daily life, however the internet of thing is still maturing "The Internet of Things allows people and things to be connected anytime, anyplace, with anything and anyone, ideally using any path/network, and any service" [5], [6] as definition given by European research cluster of IoT (IERC) The International Telecommunication Union (ITU) views IoT very similarly: "From anytime, anyplace connectivity for anyone, we will now have connectivity for anything" [7].

The ERC vision is that "the major objective for IOT is make the smarter world with IOT where smallest devices such as Smart phones, digital cameras, etc to largest environment as Cities, Transport, Health, etc are smart or self aware.Acoording to this prediction number of devices will be connected to internet for providing services/applications[32].

Various researcher are focusing on data mining technologies and its role in other area, especially in big data ,mobile application ,IOTs,Wirless networking,etc[8],The scope of data mining are grows rapidly. There are lots of data sources where data mining is required or we can say it necessary to overcome the traditional approach problems.

A´ lvaro Villalbaa, Juan Luis Pe´reza, publish a paper on data mining and IOT.they planned a platform for mining the information associated to the IoT, including both sensors data and meta-data.this platform consist two major component :First,servIoTicy for storing and processing data,it provides ability to ingest, transform on real time and query data generated by sensors;, and second iServe ,its provides capabilities to circulate, discover and use sensors meta data based on semantic information related to them [10].

Several services and applications of different technology levels coexist within the current utility grid. In this sense, it is necessary to establish techniques that provide the capability to integrate information from different architecture and technological levels. These technologies increase the robustness of the management systems related to the utility grid.

Juan I. Guerrero ∗, Antonio García, propose heterogeneous data source integration based on IEC standards and metadata mining. Additionally, an automatic data mining framework is applied to model the integrated information. In this way, this technology provides an easy-to-use and adaptive platform to integrate and model information [11].

In this paper, we survey on Data mining role in Internet of Things to handle and analyze large amount of data for improving the quality of data for services and applications. We also describe the all internal layer of data mining model for IOT. We also discuss challenges and Issues for IOT and Data mining and lastly IOT application with use of data mining in them.

Basically, Data mining means classification and analysis of data,different analysis can be apply such as classification, clustering, association analysis, time series analysis, outlier analysis, etc to extract the meaningful statistics and other characteristics of the data .

3. DATA MINING MODEL FOR THE INTERNET OF THINGS IOT

It is predictable that IOT is the future of software industry and numerous number of devices are connected with this to provide services and applications and they will produce large amount of data (or big data).As a survey 90% of data in the world today has been created in the last two year And, says the report, with new devices, sensors, and technologies emerging, the data growth rate will likely accelerate even more. [21]. So it is become an important task to store and analyze these large scale heterogeneous data efficiently to filtered and extract the information from this large data Set otherwise it can put them in an unassailable position. This is largely due to the vast increase of dark data, meaning all the unstructured data from the Internet, social media, voice and information from connected devices

Advance tools, technique and methods has been used to handle and analyze this big data, But still, Many data analysts are suggesting the digital universe will be 40 times bigger by 2020!. So, the first issue will be data which generated by sensors, social media, digital devices and mobile phone is Big Data and too hard to processed by the tools available today. The another issue As Baraniuk[15], describe in his paper because of IoT
data processing, communication, and storage capability of sensor will be changed so that the design and implementation issues of information system will be changed [18], [19], [20].

As described in [22] [23], data processing is required for big data or complex data that is generated by sensors and devicesit may help us develop friendlier or simple systems for smart city or smart home. thera are many potential applications which have been developed by analyzing the big data.information Discovery from Database is certainly one of them and it changed into effectively implemented to numerous domains to find hidden statistics out of the facts in question, it has become the initial step for numerous data structures. it's far predictable that discovery statistics from database capable of discover “something” from IoT, by using performing the subsequent steps: choice, preprocessing, transformation, information mining, and interpretation assessment [24].

As depicted in Fig. 2, IoT collects facts from diverse assets, which can also include facts for the IoT itself. KDD, whilst applied to IoT, will convert the statistics accumulated with the aid of IoT into beneficial records which can then be modified into knowledge. The statistics mining step is responsible for extracting patterns from the output of the facts processing step after which feeding them into the selection making step, which takes care of reworking its enter into beneficial information. it's far vital to notice that everyone the stairs of the KDD method can also have a strong effect on the very last effects of mining. for instance, no longer all of the attributes of the statistics are beneficial for mining; so, feature choice is normally used to select the key attributes from each record within the database for mining. The consequence is that information mining algorithms may additionally have a difficult time to locate beneficial facts (e.g., setting styles into appropriate organizations) if the selected attributes can't absolutely represent the characteristics of the information. it's also important to note that the data fusion, large scale statistics, facts transmission, and decentralized computing issues may additionally have a stronger effect on the machine overall performance and provider satisfactory of IoT than KDD or facts mining algorithms on my own might also have at the conventional packages [24].

As shown in figure 2, IoT collects data from various

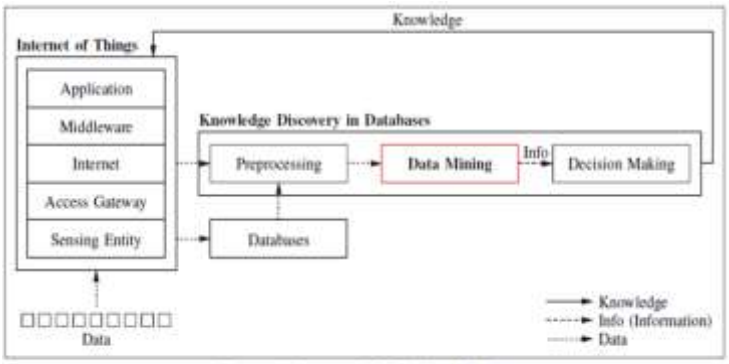

Figure 2 : The architecture of IoT with KDD.

Multi layer data mining model

According to the survey of big data mining system and IoT architecture, we endorse the multilayer architecture for IoT and large data mining device. on this system , it includes 5 layers, software and hardware resource layer,Data collection layer,data management layer ,data processing layer and service layer as shown in Figure 7
The advised massive information mining system is

 software and hardware resource layer : its include various software and hardware IoT devices to collect various smart object's data, which are RFID stream data, GPS data, satellite data, positional data and sensor data etc, such as sensors, RFID, cameras and other devices.

 Data Collection layer: In the big data mining system, structured data, semi-structured data and unstructured data can be integrated.

 Data Processing Layer: this layer integrates the data, time and other factors, so it provides a high-level mechanism for data processing of IoT [37]. Lots of open source solutions are integrated, including Hadoop, HDFS, Storm, Oozie,etc. data processing layer is used to analyze data in IoT effectively. Then we can aggregate, organize and analyze data according to the requirement [25].

 Service: Data mining service layer is depends on data management and data processing. Data mining functions will be providing as service. Various object-based or event-based data mining services, such as classification, forecasting, clustering, outlier detection, association analysis or patterns mining, are provided for applications, e.g., supply chain management, inventory management and optimization etc[25].

 Security/Privacy/Standard: Security, privacy and standard are very important to big data mining system. Security and privacy protect the data from unauthorized access and privacy disclosure. Big data mining system standard make data integration, sharing and mining more open to third part of developer[25][26].

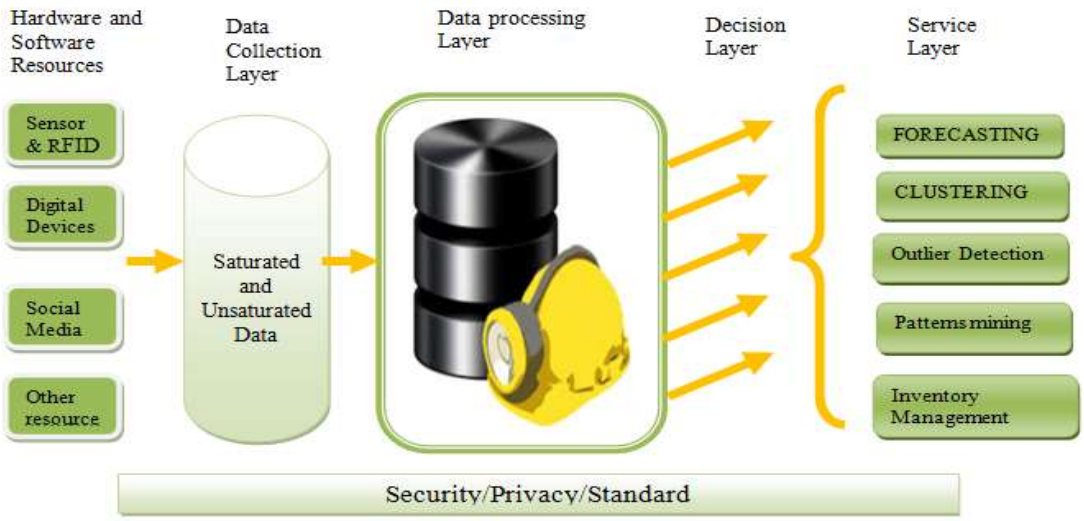

Figure 3: Multi-layer data mining model

The main objective of big data mining model is to develop an efficient structure to handle vast amount of data that is generated by IOTs and mining that data to extract useful information. Following key point we need to be considering during design of data mining framework, security & privacy of data, the data sharing mechanism, the growth rate of data, etc. To create a well designed data mining framework for IOT is a very important direction and a big challenge.

Data mining is the process of knowledge discovery by analyzing the data store in large repository. The field of data mining has been growing rapidly due to its broad applicability, achievements and scientific progress,

4. IOT AND DATA MINING APPLICATION

The internet of things is the new name given to future internet. There are two different subsets, the "consumer IOT" that includes wearable computers, smart household devices and network appliances, another one is "Industrial IOT" includes networked smart grid, manufacturing, medical and transportation [31].Some IOT application are addressing community needs and advancement to enabling technologies such as nano electronics and cyber physical systems continue to be challenged by a variety of technical, institutional and economical issues [38]. Here are some of the IoT smart application [34] discusses in this section.

4.1. Smart cities - eight smart features or sub application will lead the evolution of smart cities, which include Smart economy, smart building, smart mobility, smart energy, smart information communication and technology, smart planning, smart citizen and smart governance. It is assume that by 2015, there will be about 40 smart cities globally in Europe, the largest smart city initiatives completely focused on IOT. This smart city project aims at deploying an IoT infrastructure by interconnected sub-applications or services to establish a large-scale interconnected heterogeneous network for IoT/CPS applications, with the aim of achieving the best use of public resources in cities [32][12].
In public service area, data mining can be used to discover public needs, improve service performance, decision making with automated systems to decrease risks.

4.2 Smart Medical - smart health platform require monitoring devices, characterized by application-specific solutions that are mutually non-interoperable and are made up of diverse architectures [12]. These application require gathering of data from sensors, these application also required network connectivity for access to infrastructureral services to support interfaces and displays. So data mining is essential in healthcare .Heterogeneous medical data have been generated in various healthcare organizations in various formats such as, payers, medicine providers, pharmaceuticals information, prescription information, doctor's notes or clinical records which are creating day by day [26]. These quantitative data can be used to do clinical text mining, predictive modeling , survival analysis, patient similarity analysis and clustering, to improve care treatment and reduce waste [26].Internet of things application have future market potential for electronics health service or a new type of health industry[12][31][32].

Figure 4 : IOT Applications

4.3 Industrial IOT: the industrial IOT will eventually have much larger economic impact. The IIOT will bring entirely new infrastructure to our most critical and impactful societal system. The opportunity to build truly intelligent distributed machines that can greatly improve function and efficiency across virtually all industries is indisputable. The IIOT is the strategic future of most large companies, even traditional industrial manufacturers and infrastructures providers [31]. Data mining also perform major role to making smart industries such as retail, banking, and tele-communications, classification, clustering can be applied to this area. The main objective of IIoT will control expensive mission critical systems, unlike connecting consumer devices [32] [12].

4.4 Smart energy and smart grid - In the smart gird, distributed energy generators are introduced to improve the utilization of distributed energy resources, electric vehicles are introduced to improve the capability of energy storage and reduce emission of *CO*2, and smart meters and bidirectional communication networks are introduced to achieve the interactions between customers and utility providers. With these techniques, the smart grid can achieve great reliability, efficiency, safety, and interactivity [29].To developed a smart grid for replacing the traditional power grid, require integrating IOT with cyber physical system. In this respect the internet of energy concept is defined as a network infrastructure based on standard and interoperable communication transceivers, gateways and protocols that will allow a real time balance between the local and global generation and storage capability with energy demand [32]. Sophisticated and flexible data filtering, data mining and processing procedures

and system will become necessary to handle the high amount of raw data provided by billions of data sources. Smart power grid has to be able to react correctly and quickly to fluctuations in the supply of electricity from energy sources such as wind and solar facilities [31] [32].

4.5 Smart transportation – Smart transportation means adding new functionality to individuals and /or making transport easier and safer. The internet connection with vehicles represents future trends for smart transportation and mobility applications[12].we can also used IOT in traffic management and control system, where self smart vehicles able to organize themselves in order to avoided traffic jams and to optimize drive energy usage[32]. Smart sensors in the road and traffic control infrasture need to collect information about road and traffic status, weather condition, etc. this require robust sensors and effective data mining technique to classified and filtered the smart data, which are able to reliably deliver information to the system. Smart transportation also required safe and secure communication with elements at the network edge, inter vehicle communication and vehicles to infrastructure communication [29] [32].

5. CONCLUSION

To discover useful knowledge from real-world and simulated big data, business intelligence professionals and data scientists face with new challenges. In today smart era, the conventional technologies and methods cannot store and analyze a large amount of data. Since the "Internet of things" gain in popularity, the attention is focused on the development of new data mining model for internet of things with a possibility to access high-performance computing environments. In This paper, review of the existing data mining software and solutions with internet of thing, has been drawn. Smart environment features and the new technologies related to information management are the future of the new smart services and applications. Several services and applications of different technological levels coexist within the current utility grid. In this sense, it is necessary to establish techniques that provide the capability to integrate information from different architecture and technological levels. These technologies increase the robustness of the management systems related to the smart environment. We also discuss the challenges and issues of data mining towards IOT, and various applications of internet of things. We believe that further research is needed to construct efficient models for such applications the models could be improved by adding new information, and performing the modeling algorithm. Future research lines include the application of this technology to other types of database, such as document-based and key-value databases

REFERENCES

[1] Anne James, Joshua Cooper, Keith Jeffery, and Gunter Saake "Research Directions in Database Architectures for the Internet of Things: A Communication of the First International Workshop on Database Architectures for the internet of things (DAIT 2009),"BNCOD 2009: 225-233.
[2] 1. Luigi A., Antonio I., Giacomo M. 2010.The Internet of Things: A survey. Science Direct Journal of Computer Networks, Volume 54, Pages: 2787–2805.
[3] S. Haller, S. Karnouskos, and C. Schroth, "The Internet of Things in an enterprise context," Future Internet Systems (FIS) , LCNS, vol. 5468. Springer, 2008, pp. 14-8.
[4] Cooper J, James "A. Challenges for Database Management in the Internet of Things," IETE Tech Rev. 2009. 26:320-9.
[5] P. F. Harald Sundmaeker, P. Guillemin, and S. Woelfflé,Vision and Challenges for Realising the Internet of Things .Pub.Office EU, 2010
[6] V. Ovidiu and F. Peter, Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems , 2013.
[7] IT Union, "ITU Internet report 2005: The Internet of Things," 2005.
[8] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, "CAP: crowd activity prediction based on big data analysis," IEEE Network, vol. 28, no. 4, pp. 52–57, 2014.
[9] J. Wan, D. Zhang, Y. Sun, K. Lin, C. Zou, and H. Cai, "VCMIA:a novel architecture for integrating vehicular cyber-physical systems and mobile cloud computing," Mobile Networks and Applications, vol. 19, no. 2, pp. 153–160, 2014.
[10] A´ lvaro Villalbaa, Juan Luis Pe´reza, David Carreraa, Carlos Pedrinacib, Luca Panzierab servIoTicy and Serve: a Scalable Platform for Mining the IoT Procedia Computer Science Volume 52, 2015, Pages 1022-1027
[11] Juan I. Guerrero, Antonio García, Enrique Personal, Joaquín Luque, Carlos León "Heterogeneous data source integration for smart grid ecosystems based on metadata mining" Elsevier Volume 79, 15 August 2017, Pages 254–268.
[12] Introduction to Data Mining and its Applications by S. Sumathi, S.N. Sivanandam ol. 20. Wojciech Penczek, Agata Pórola Advances in Verification of Time Petri Nets and Timed Automata, 2006 ISBN 3-540-32869-6 [13] N. Ali and M. Abu-Elkheir, "Data management for the internet ofthings: Green directions," inProc. IEEE Globecom Workshops, 2012,pp. 386–390.
[14] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang "Data Mining for Internet of Things: A Survey" IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 16, NO. 1, FIRST QUARTER 2014
[15]R. G. Baraniuk, "More is less: Signal processing and the data deluge,"Science, vol. 331, no. 6018, pp. 717–719, 2011.
[16] D. Reed, D. Gannon, and J. Larus, "Imagining the future: Thoughts on computing,"Computer, vol. 45, no. 1, pp. 25–30, 2012.
[17] M. Hilbert and P. L´opez, "The world's technological capacity to store, communicate, and computer information,"Science, vol. 332, no. 6025,pp. 60–65, 2011.
[18] S. Madden, "From databases to big data," IEEE Internet Computing,vol. 16, no. 3, pp. 4–6, 2012.

[19] Y. K. Chen, "Challenges and opportunities of internet of things," in Proc. Asia and South Pacific Design Automation Conference, 2012,pp. 383–388.
[20] D. Bandyopadhyay and J. Sen, "Internet of things: Applications and challenges in technology and standardization," Wireless Personal Communications, vol. 58, no. 1, pp. 49–69, 2011.
[21] V. Mayer-Schonberger and K. Cukier, Big Data: A Revolution ThatWill Transform How We Live, Work, and Think. Boston: Houghton Mifflin Harcourt, 2013.
22] Libelium Comunicaciones Distribuidas S.L., "50 sensor applications for a smarter world," 2012, available at http://www.libelium.com/top 50 iot sensor applications ranking.
[23] R. Macmanus, "Top 10 internet of thingsproducts of 2009," 2009, available at http://www.readwriteweb.com/archives/top 10 internet of things products of 2009.php.
[24] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Magazine, vol. 17, no. 3, pp.37–54, 1996.
[25]Shen Bin, Liu Yuan, Wang Xiaoyi "Research on Data Mining Models for the Internet of Things" 978-1-4244-5555-3/10/$26.00 ©2010 IEEE
[26] Data Mining for the Internet of Things: Literature Review and Challenges, ACM digital library Volume 2015, January 2015 Article No. 12
[27] Ming –Syan chang, Jiawel Han, Philip S. yu "data Mining: an overview from database perspective "IEEE Transaction on knowledge and data engineering vol 8,no 6 December 1995.
[28] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang " Data Mining for Internet of Things: A Survey" IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 16, NO. 1, FIRST QUARTER 2014 77
[29] Jie Lin_, Wei Yuy, Nan Zhangz, Xinyu Yang_, Hanlin Zhangx, and Wei Zhao "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, andApplications" 2327-4662 (c) 2016 IEEE internet of things journals
[30] Rolf H. Weber "Internet of Things – New security and privacy challenges" Elsevier science direct available at www.sciencedirect.com
[31] Internet of Things and Data Analytics Handbook edited by Hwaiyu Geng
[32] Ovidiu Vermesan & Peter Friess "Internet of Thjings –Converging Technologies for smart Environment and integrated Ecosystem" , River Publication ,ISBN 978-87-93379-20-6
[33]]C. Hsinchun, W. Chung, J. J. Xu, G. Wang, Y. Qin, M. Chau,"Crime data mining: a general framework and some examples", Computer, 37(4), pp.50-56, 2004.
[34] Mohammad Abdur Razzaque, Member, IEEE, Marija Milojevic-Jevric, Andrei Palade, and Siobhán Clarke "Middleware for Internet of Things: A Survey" IEEE INTERNET OF THINGS JOURNAL, VOL. 3, NO. 1, FEBRUARY 2016
[35] Anne H. H. Ngu, Mario Gutierrez, Vangelis Metsis, Surya Nepal, and Quan Z. Sheng, Member, IEEE "IoT Middleware: A Survey on Issues and Enabling Technologies" IEEE INTERNET OF THINGS JOURNAL, VOL. X, NO. X, MARCH 2016
[36] Ivan Kholod ,Mikhail Kuprianov ,Ilya Petukhov "Distributed data mining based on actors for Internet of Things" IEEE, Embedded Computing (MECO), 2016 5th Mediterranean Conference on 01 August 2016 DOI: 10.1109/MECO.2016.7525698
[37] Farheen Fatima1, BatulNaeem Husain2, Mohammed Azharuddin3, MohammedAbdul Mabood "Internet of things:A Survey on Architecture, Applications, Security, Enabling Technologies, Advantages & Disadvantages "International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015Copyright to IJARCCE DOI 10.17148/IJARCCE.2015.412117
[38] O. Kurasova, V. Marcinkevičius, V. Medvedev, A. Rapečka, P. Stefanovič "Data Mining in Big Data Era: Methods and Technologies" 6th International Workshop DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS https://www.mii.lt/datamss/files/liks_mii_drusk_2014_abstract_last_print_.pdf?
[39] https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/