

# Internet of Things with the Perspective of Data Analytics and Machine Learning

Dr. Sumita U. Sharma,  
Deptt. Of Electronics,  
DCPE, HVPM, Amravati

**Abstract:** The Internet, as we know today has become the most essential commodity of our daily life. The conventional form of communication that happens on the internet is the H2H (Human to Human) or H2M (Human to Machine) which is very common when users access the internet. A new type of communication called M2M (Machine to Machine) has become a trend and a buzzword which is used in the Internet of Things. Through IoT, the real world “things” or objects transform into virtual intelligent objects. IoT and big data are the two most talked about and converged technologies in the modern times. Data analytics and the machine learning are also very prevailing tools that are dominant in the recent years. This paper focusses on the trends in the present day and the future of IoT with data analytics and machine learning.

**Keywords:** *Big Data, Internet of Things (IoT), Data Science, Data Mining, Machine Learning*

\*\*\*\*\*

## I. INTRODUCTION

Internet of Things or IoT in short, is the modern technological revolution and a paradigm shift from the conventional H2H (Human to Human) or H2M (Human to Machine) communication to things to things or M2M (Machine to Machine) communication [1]. It is stated that the IoT is the future of computing and communication systems used in various innovative fields like wireless sensor networks, RFID systems and nanotechnology. The concept of Internet of Things is not apparently new and it goes back to early 80's where an innovative coke vending machine was connected to the Internet. The machine was capable of sending reports like inventory of drinks available in the machine and also report if they were cold enough [2]. In 1991, the existing vision of IoT in terms of ubiquitous computing was first assumed by Mark Weiser [3]. Bill Joy in 1999, put forth the idea about Device to Device communication in his research paper [4]. During same period, Kevin Ashton suggested the term “Internet of Things” to describe a system of interconnected devices [5].

The primary purpose of the IoT is to make independent exchange of information and data from the real world objects or devices that are interconnected through internet technologies and protocol like TCP/IP. These interconnected objects can be all around us and powered by technologies like RFID (Radio Frequency Identification) and the popular WSN (Wireless Sensor Networks) which are sensed by the sensor devices and further processed for decision making, on the basis of which an automated action can be performed. The data captured through the sensors can also be sent to a server by the IoT devices. The large volumes data collected at the server is then stored in the distributed storage and processing system like Hadoop which

is later used for performing data analysis for further decision support systems.

Yet another modern day technology in vogue is the big data which specifically means state-of-the-art techniques and technologies to capture or collect, store, distribute, manage and analyze data in large volumes usually measuring in petabytes and exabytes or even larger size. The data stored could be of different structure and formats. Four words, generally used with the big data are *variety* (data in various forms), *volume* (referred to size of data) and *velocity* (speed at which the data is added) and *veracity* (uncertain provenance). Big data can be structured, unstructured or semi-structured, and handling the data which the conventional systems are incapable of. Data can be produced and generated by large number of sources like computer systems, mobile devices, IoT devices and many more [6]. The data can arrive at different speeds and size. Massive amount of data thus gathered and stored need an efficient and inexpensive way for processing. The parallel processing or cluster computer makes it possible to process the big data.

Furthermore, to manage and process the big data in a distributed environment, a new system with advanced architecture, appropriate algorithms and techniques are needed which would help to process and analyze data to extract meaningful results and hidden knowledge which would lead to business improvement. For the purpose of storing big data, distributed processing and performing analytics the Hadoop framework is used. Hadoop is an Apache Software Foundation project and an open source software project developed in platform independent language Java. Hadoop enables the distributed processing of large data across many clusters of commodity servers. It is designed to scale up from a single server to thousands of machines, with a very high degree of fault tolerance [7].

## II. IMPORTANT DEFINITIONS

### A. *Internet of Things (IoT)*

The Internet of Things (IoT) is the inter-networking of physical devices, machines, vehicles buildings, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data and performs some action as we provided [8]. The IoT allows objects to be sensed or controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based or can say automation based systems. This results in improved efficiency, accuracy and economic benefit in different activities that helps to reduce human intervention.

### B. *Data science*

**Data science**, also known as data-driven science, is an interdisciplinary field to extract knowledge or insights from data in various forms by using scientific methods, processes, and systems. It comes into the scene along with the generation of Big data. It has the powerful techniques that extract knowledge from structured or unstructured type of data and performs similar functionality like data mining [9]. Data Science is the scientific concept, data science can be called as "concept to unify statistics, data analysis and their related methods" that helps to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many different fields and covering the areas of mathematics, statistics, information science, and computer science in broad scene. In particular these relates from the subdomains of machine learning, classification, cluster analysis, data mining, databases, visualization, etc. Mainly it relates with the large amount of data that is generating today and their processing techniques.

### C. *Machine learning*

Machine learning is a multidisciplinary domain whose primary focus is on the theory, performance, and properties of learning systems and its algorithms. It is a greatly interdisciplinary area that builds upon the ideas from many diverse fields like artificial intelligence, statistics and statistical analysis, optimization theories, information theory, cognitive science, optimal control, and various disciplines of science, mathematics and engineering. It has been implemented in broad range of applications and the machine learning is said to be covered almost all scientific domains and areas [10]. Machine learning has been applied to a large variety of problems including informatics, autonomous control systems, cognitive systems, and data mining. In general, the machine learning techniques are categorized into three sub types as supervised learning, unsupervised learning, and reinforcement learning. The supervised learning needs training with labeled data which has inputs and desired

outputs. On the other hand, the unsupervised learning does not need labeled training data but the environment itself provides inputs without desired targets. The reinforcement method is capable of learning from feedback received from the interactions with an external environment. These three important types of learning paradigms become basis for many types of theory mechanisms and application services that have been proposed for dealing with data tasks.

## IV. THE COMBINATION OF IOT AND BIG DATA ANALYTICS

The intersection of the IoT and big data is a multi-disciplinary field. Highly specialized skills are essential if businesses need to extract maximum value from it. Two kinds of resources will be required business analysts who can structure appropriate questions to be asked of the available data and present the results to decision makers; and data scientists who can organize the rapidly evolving cast of analytical tools and curate the veracity of the data entering the analysis pool. In rare cases, the business analyst and the data scientist may be one and the same valuable person.

Our increasingly connected world now-a-days moving towards automation by combining with low-cost sensors and distributed intelligence. This will transform the data generation and management process and creates the impact on industry, producing more data than humans will be able to process. Now with the today's need it is necessary to adapt and evolve quickly to maintain their place in the competitive landscape. And it becomes necessary to make sense and understand benefits from these new sources of information and intelligence embedded in our environment.

Basically, Internet of Things (IoT) is set to change how we work and interact with the world in every possible manner. A few scenarios and possibilities enabled by this internet of things that helps to get the idea about the IoT [11].

### *End user applications:*

- As a person approaches the door of your home, it senses his presence and automatically unlocks itself. As soon as he close it to go out of home, it asks all the energy units in the house like lighting, heater, stove, refrigerator to go into energy preservation mode. When you come back, the reverse action will happens.
- The wrist band has come that senses when user fall asleep and automatically asks his headphones and lights in his room to switch off.
- The sensors in the soil of garden gauge the level of moisture in soil and accordingly switch the watering unit.
- In Health care – Imagine a person's watch is monitoring his heart rate regularly and notifies him

at the first instance of any irregularity. It can also send a message to his doctor or to nearby hospital.

#### **Commercial applications:**

- If the train and its track loaded with sensors, which continually monitor their wear and tear. Even before the train reaches its destination, you would know the repairs and the changes required. Similarly, you do not depend on the generic guidelines about the life of tracks to replace them. The same concept can be applied any kind of machineries that can be connected with sensors, to aircraft engines, wind turbines, etc.
- The important concept regarding residence i.e. of Smart cities can become possible with the help of IoT. It consists of designing a city, infrastructure locations according to complete plan. Width and proper locations of the roads. Designing alternative routes to ease traffic. Correct location for a hospital, schools, and commercial areas along with their capacity. All these decisions being made on data and with planning and using Machine learning application with IoT.

#### **IV. INSIGHTS OF DATA ANALYTICS AND MACHINE LEARNING WITH IOT**

The recent trends show that the big data and big data analytics are commonly used to describe the data sets and data analytical techniques in applications that are very large ranging from terabytes to exabytes and densely complex data generated from sensor networks, RFID devices and the social media that they require unconventional, powerful and unique data storage, management, analysis, and visualization technologies. The volume, velocity, variety and veracity (used to describe big data), IoT generated data makes it no easy task to select or build an analytics solution that can generate useful business insight and provide the business intelligence which leads to the decision support systems.

#### **Challenges to the Internet of things:**

There are a few significant challenges that need to be overcome in order to make smooth use of this necessary technologies [11]:

- As IoT is based on M2M communication, one of the main problems is to get machines from different manufacturers to talk to each other. It creates a problem while communicating sensitive customer data across to each other.
- While the applications are immense, so are the scenarios to use these machines. Each scenario requires a different style of communication –

different latency, different frequency, and different length of data. For example, a sensor measuring temperature of human body needs to communicate the data once in a while. The driverless cars need to be on their toes every single second [12].

- Internet of things will expose whatever privacy is left today to be exposed – both at personal and industrial levels. The proponents of customer privacy would be shouting on top of their voices by now.
- One of the main challenges that each device has to face is of information security. Each of these devices would have their own security challenges. As we all know, our PCs, Laptops and mobiles, even after years of efforts still have security challenges. And as the use of internet and the devices connected to the internet goes on increasing, this challenge now goes to a different level.

#### **Resolving Challenges by:**

- Understand the relative maturity of enterprise capabilities in the realms of product technology and IT;
- Understand the types of IoT functionality that can be incorporated and where new capabilities will impact customer value;
- Understand the role of machine learning and predictive analytics models; and rethinking business models and value chains based on how quickly the market is changing and the relative agility of competitors.
- Understanding Product and IT Maturity as Mining equipment is technologically complex but has longer equipment life cycles and relatively slower evolution than scientific research instrumentation. Another factor needs to be considered - that of IT process maturity. Each type of organization would benefit from IoT enablement; however, the models for that evolution will vary.
- The implication for the scientific instrumentation firm might be that IoT will allow for functionality updates of field instrumentation, but the firm might not want to attempt to optimize a laboratory information ecosystem consisting of multiple classes of equipment.

#### **Form data science perspective:**

As the use of internet and online applications is growing on increasing day-by-day and as IoT comes into picture it stores the data from every second with the use of sensors. This is how the data generated through internet of things is expected to increase in gross amount. Cisco expects the data emitted by all mobile devices to grow 4 times in 3 years! [13]. Now,

it is necessary to imagine the hugeness and importance for processing the big data from data science professionals. Some of them are as follows:

- First and foremost, we will need better and more agile infrastructure as the data that seems big for today would become small for tomorrow.
- It becomes difficult to marry the data across devices and make sense out of it. Some possibilities that one can develop with the integration of IoT with data science and machine learning are develop smart watch and mobile to predict health conditions of a common man, could inform the doctors in advance with notifications, predict the number of doctors required to be on duty day on day, etc. In simple words, with this explosion in data, it is necessary to handle this data and make more sense of it.

### V. CONCLUSION

The concept of Internet of Things (IoT) arises from the need to manage, automate, and explore all devices, instruments, and sensors in the world. In order to make wise decisions both for people and for the things in IoT, data science methods are integrated with IoT technologies for handling huge amount of data that is generating on daily basis. Along with these machine learning techniques are involved that helps in discovering novel, interesting, and potentially useful patterns from data and applying algorithms to the extraction of hidden information. This paper performs the survey of insights of IoT with Data science and Machine learning. It will show the large amount of application that are working and can be developed on the concepts of IoT that are necessary in today's world. Based on the survey of the current research, it can be said that there are immense possibilities that one can develop with the integration of IoT with data science and machine learning.

### REFERENCES

- [1] Somayya Madakam, R. Ramaswamy, Siddharth Tripathi, Journal of Computer and Communications, 2015, 3, 164-173 Published Online May 2015 in SciRes.
- [2] Guicheng Shen and Bingwu Liu, "The visions, technologies, applications and security issues of Internet of Things," in E-Business and E-Government (ICEE), 2011, pp. 1-4
- [3] Ling-yuan Zeng, "A Security Framework for Internet of Things Based on 4G Communication," in Computer Science and Network Technology (ICCSNT), 2012, pp. 1715-1718
- [4] Jason Pontin, "Bill Joy's Six Webs," MIT Technology Review, 29 September 2005.
- [5] Kevin Ashton, "That 'Internet of Things' Thing", RFID Journal, 22 June 2009
- [6] Harshwardhan S. Bhosale<sup>1</sup>, Prof. Devendra P. Gadekar, A Review Paper on Big Data and Hadoop, International

- [7] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu, Shuo Feng, A survey of machine learning for big data processing, EURASIP Journal on Advances in Signal Processing, December 2016, 2016:67
- [8] "Internet of Things: Science Fiction or Business Fact?" , Harvard Business Review. November 2014. Retrieved 23 October 2016.
- [9] Dhar, V. (2013). "Data science and prediction". Communications of the ACM. 56 (12): 64. doi:10.1145/2500499.
- [10] Alpaydin, Ethem (2010). Introduction to Machine Learning. London: The MIT Press. ISBN 978-0-262-01243-0. Retrieved 4 February 2017.
- [11] "Internet of Things (IoT) and its impact on data science", [Online] available: <https://www.analyticsvidhya.com/blog/2015/04/internet-of-things-impact-data-science/>
- [12] Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos, and Xiaohui Rong, "Data Mining for the Internet of Things: Literature Review and Challenges", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks, Volume 2015, Article ID 431047.
- [13] Yue Xu, Prof. Raj Jain, "Recent Machine Learning Applications to Internet of Things (IoT)", [Online] available: [http://www.cse.wustl.edu/~jain/cse570-15/ftp/iot\\_ml/index.html](http://www.cse.wustl.edu/~jain/cse570-15/ftp/iot_ml/index.html).