Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 1, pp. 23–33, DOI 10.12694/scpe.v23i1.1969

INFORMATION RETRIEVAL AND DATA ANALYTICS IN INTERNET OF THINGS: CURRENT PERSPECTIVE, APPLICATIONS AND CHALLENGES

KRUTI LAVINGIA* AND RACHANA MEHTA†

Abstract. The Internet of Things has emerged as an evolving paradigm and has developed its presence in a variety of domains around us. The emergence of IoT has also emphasized the need to cater to challenges such as interoperability, smart IoT components adoption, authentication and authorization, networking, information retrieval, and several other issues. The ubiquitous nature and interconnection between various devices supported by machine learning, artificial intelligence, cloud computing, big data, and blockchain lead to a generation of large amounts of data. In order to find useful information from such data is a tedious task and involves high computation. The domain of Information Retrieval helps us to identify and manage environmental factors of data collected through sensors. The data gathered may be heterogeneous and from different sources. This demands the need for better retrieval efficiency, accuracy, and systematic models for gathering and managing the sensed data. Designing such a model with security and privacy is a major concern. The acquired knowledge from those models will be helpful for data analytics, performance-boosting, decision making, and managing the resources efficiently. A detailed study of the importance of Information Retrieval and Data Analytics in the Internet of Things is presented in this paper.

Key words: Internet of Things, Data analytics, Information retrieval, Big data, Data mining, Smart City, Machine Learning

AMS subject classifications. 68P20

1. Introduction. The Internet of Things (IoT) has emerged exponentially in the past decade owing to the advances in wired and wireless devices in and around us [1]. Internet of Things involves the interaction and intercommunication among the connected heterogeneous devices [2]. The communication among those devices takes place using technologies such as Bluetooth, Global System for Mobile Communication (GSM), ZigBee, Radio Frequency Identification (RFID) sensors and actuators, WiFi, and several others. These technologies allow phones, laptops, gaming consoles, sensors to communicate with one another and exchange information [1, 2].

The impact of IoT technology on the Sustainable Development Goals (SDGs), devised by the United Nations (UN) has been a significant area of research. The current applications of IoT such as Smart Cities, Smart Learning, Smart Agriculture, Smart Healthcare, IoT based Intelligent Systems for Education, Water Management, Smart Grid, IoT in smart manufacturing, industrial automation are directly in line with the 2030 sustainability goals of the UN. A project initiative is taken by the World Economic Forum for the IoT for Sustainable Development, more than 640 projects are being analyzed for their applicability. The main focus is on developing scalable and reusable models which can be used by different stakeholders [7]. This signifies the need and importance of IoT-based solutions and their use.

IoT is prevalent and making its dominance in almost every field around us like home automation, logistics, manufacturing, healthcare, agriculture, smart grid, smart city, and much more [1, 3, 4, 5]. The wide adoption of IoT, the Internet, and the Industrial revolution like Web 4.0 has made a significant impact on the working culture and digitization. This has led to many companies going digital and leveraging the benefits of the online environment like working from anywhere, working at any time, accessing data from anywhere, increasing operational efficiency, reduction in infrastructural cost, and much more. Especially in the current scenario of COVID-19, this digitization has led the companies and industries to go and work smoothly. The boom in digitization and advancement of IoT has also floated up the challenges coming with it. Some of them are interoperability among the connected devices, adoption of the smartness among the IoT components, security

*CSE Department, Institute of Technology, Nirma University, India (kruti.lavingia@nirmauni.ac.in).
†CSE Department, Institute of Technology, Nirma University, India (rachana.mehta@nirmauni.ac.in).

and privacy concerns of authentication and authorization, issues pertaining to network and communication, retrieval of information, and analytics on data [1, 3, 6, 9].

The ubiquitous behavior and interconnection among various homogeneous and heterogeneous devices supported by new edge technologies like Artificial Intelligence, Machine Learning, Deep Learning, Blockchain, and Cloud Computing have led to a generation of large amounts of data. In order to fetch meaningful information from such a heap is a complex and time-crunching task, as well as involves a high degree of computation and resource utilization. This process of finding out useful information is known as Information Retrieval (IR). In the context of IoT, the IoT devices and sensors are the primary sources of data generation, this data may come in varied forms and changes according to the application at hand. Since the IoT is based on the intercommunication of networked devices, the IR task also gets affected by the network technology being used for a particular application. This brings in the new concern of efficient retrieval of data, devising the models that can gather sensor data, record data, and manage them using appropriate indexing mechanisms, addressing the security and ethical issues, and extracting information in real-time. On the other hand, the task of data analytics involves the work on data that is collected or retrieved. Data analytics perform various operations such as data cleaning, data preprocessing, data mining to extract the trends, patterns, hidden features, and other useful information present in the collected data. The analytics helps to know the working pattern for businesses, trends, and patterns which are, unless invisible, help to understand the dynamics of the working model and efficiently leverage them for futuristic work. Data Analytics in IoT follows different mechanisms and techniques for different applications. The task of data analytics is dependent on data used in the application, data generated in the application, and the approach used for analytics purposes. The current paper gives insights on the current work done in Information Retrieval and Data Analytics in IoT along with their issues and possible solutions.

The remaining part is divided as follows: Background provides the insights on IoT environment and what data analytics is, the next section gives a detailed overview of various use cases of IoT-based applications using data analytics, and the challenges pertaining to the data analytics are considered, followed by conclusions.

2. Background. Before we dwell on the concerns of Data Analytics and Information Retrieval using IoT, this section gives an understanding of the terminologies on hand.

2.1. IoT Environment. The Internet of Things involves the interconnection of various sensor-based devices and intercommunication among them with the help of varied communication technologies. The general IoT environment consists of mainly IoT devices, and networking technologies [8]. The IoT devices consist of sensors and actuators directly used or embedded within the system. Commonly used sensors include a temperature sensor, humidity sensor, pressure sensor, proximity sensor, infrared sensor, gas sensor, accelerometer, level sensor, gyroscope, motion sensor, chemical sensor, and much more. A single IoT device may consist of more than one sensor to measure multiple factors at a single point in time. The sensor and actuator-based devices are the primary sources of data collection. The data generated through sensors and actuators get stored locally either on the IoT device itself or on the external storage of the cloud, user system, or edge. The storage is purely dependent on the IoT environment. The data stored is further useful to the end-user. The same is transmitted to the end-user or shown with the help of the user interface. The IoT sensors are embedded, and along with it, the support for network communication technologies is done using the ports and antennas. The networking and communication technologies in the IoT environment involve support for the intercommunication among the IoT devices and from IoT device to end system. In most cases, the analysis of generated data is carried out to infer the notable trends, patterns, and features.

2.2. Data Analytics. The domain of data analytics involves the analysis of data to study the environment from where data is coming, the behaviour and working of the environment, to infer meaningful relationships, to make predictions, and to decide the future actions.The data generated from the sensor devices are taken as input for the data analysis and the outcomes of the analysis are patterns, hidden features, new information, and statistics of the system which helps in decision making. The volume of the data generated from IoT devices has increased manifolds owing to the increase in usage of smart devices, the Internet, advancements in communication networks, and IoT technologies in all domains. These involve using efficient data analysis tools which support heavy data, such as big data. The task of data analytics is not only limited to using the commer-

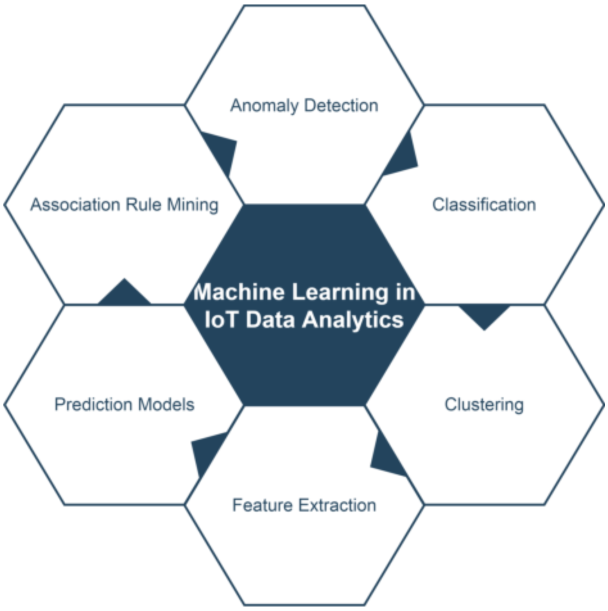

Fig. 2.1. *Machine Learning Applications in IoT*

cial data analytics software, but also includes using current edge technologies like Artificial Intelligence (AI). Machine Learning (ML), a sub field of AI and Deep Learning (DL), a subfield of ML are also prevelantly used for Data Analytics. These models use techniques like clustering, classification, prediction models, association rule mining, feature extraction, anomaly detection and many others for data analysis purpose. The current research focuses on the various ways data analytics can be carried out including, the analytics tools and various computing paradigms of AI, ML, and DL.

Fig 2.1 shows the various ML-based algorithms used for data analysis in IoT.

3. Use cases of IoT Data Analytics. There are a number of domain-specific applications in IoT that deal with huge amounts of data. For any such application, sensors are the key elements responsible for gathering information from the environment. The sensor-generated data is key to the data analysis in IoT applications. In this section of the paper, various IoT use cases for data analytics are discussed, provided that are helpful for the sustainability perspective.

3.1. Smart city. The smart city is one of the highly explored and leveraged IoT applications owing to its enormous benefits. The Smart city application covers all the IoT domains. It includes various IoT use cases like smart traffic management, smart waste management, smart governance, smart people, smart infrastructure, smart grid, smart people, smart healthcare, smart sanitation, smart monitoring, smart environment, smart economy, and much more.

Rathore et al. [10] have proposed a prototype of Smart Digital City along with the big data processing on a real-time basis. Apache Spark over Hadoop has been taken into consideration for data analysis on smart city applications like home, parking, climate, pollution monitoring, and vehicle network. The analyzed data is used by the government and municipalities for making decisions such as traffic analysis, diverting the traffic when there is congestion, showing the empty parking spots in a city, city planning based on the pollution status in a given area, and managing the water resource using the smart home data [10]. Jinping Chang et al [11] propose an adaptive heuristic mathematical model for traffic congestion. The model work includes monitoring the city traffic through sensors in all day and all-out emissions. The collected data is further sent to the central place which organizes the traffic police efficiently and sustainably. As well the traffic can be diverted to other paths, leading to reduction of carbon emission at specific places. The proposed model includes video monitoring, surveillance, GPS location tracking system, and GIS to generate the data. The data is further analyzed using the traffic status, alerts, and graph to divert the traffic and manage it appropriately. M. Ashwin et al [12] have devised an automated intelligent smart bin to resolve the waste management issue. Ultrasonic sensor and servo motor are used to detect the human presence and overall work of the bin respectively. Furthermore, the capacitive sensor is used to segregate the dry and wet waste. As well it works on the solar panel to run the bin efficiently without any external energy source. An optimal route selection algorithm is used to dispose of the bin and install the empty one.

3.2. Smart Building. The smart building stands at the core of the infrastructural aspect of a smart city. Smart buildings are useful for the effective management of a building, its construction, monitoring the progress of residents and their concerns, home automation, real-time alert management, intrusion detection in surveillance of the building, efficient utilization of energy, waste management, individual authorization, etc. Smart buildings are concerned with using the IoT at every level from architectural planning, construction to implementation and final results.

M.Dey et al [13] have proposed a novel feature extraction technique for extracting the power and temperature information from the high dimensional terminal unit data of Heating Ventilation and Air-Conditioning (HVAC). An unsupervised X-means clustering method has been applied to find the faulty unit and in the second stage, Multi-Class Support Vector Machine (MC-SVM) algorithm has been applied to find the pattern of fault detection and provide an appropriate diagnosis. This helps in reducing the amount of energy wastage. Also, an automated alert is employed to inform the maintenance team. Isaac et. al. [14] have proposed the HEMS-IoT, a framework for energy management in smart homes, comfort, and safety. The collected sensor data is analyzed using the J48 classification algorithm and WEKA tool. The analysis provides insights on the user behavior and the energy consumption by the home devices. Further RuleML and Apache Mahout have been used to provide energy saving and safety recommendations to the user. Wei Zhang et al [15] have devised the thermal comfort model for smart buildings. Machine Learning-based approach is used to design the model, it helps in bridging the gap between the controllable parameters of building and thermal comfort. Neural Network-based models have been found efficient for the data analysis and control parameter tuning using the real-time data for time, date, and weather. M.R. Bashir [16] has proposed an integrated IoT big data analytics framework for the real-time analysis of oxygen levels, luminosity, and hazardous gas in different parts of the building. The IoT sensors incorporated include the oxygen sensor, gas detector, and luminosity sensor. The real-time data generated is fed to the PySpark for the analysis purpose, and Cloudera for visualization purposes, which alerts for turning on the oxygen pump, when the oxygen level drops, and similarly for turning off the pump when it is within a threshold value. The data generated from the gas sensor is used for alerting and tuning the fire alarm. A luminosity level sensor is used to turn on and turn off the lights. This helps in the efficient management of energy and smart control of buildings.

3.3. Smart Agriculture. The Internet of Things helps in dealing with a number of challenges related to practical farming. The modernization in the field of Agriculture is taking place due to the advancements in the IoT systems. A variety of agriculture-related issues such as identification of temperature and climate beneficial for the crop, temperature, and productivity of the soil, water level required, horticulture, usage of pesticides, manure requirement, etc can be identified using different sensors and these tasks can be automated using IoT sensors and applications. Lavanya G. et al.[17] have designed a novel sensor called the NPK Sensor (Nitrogen, Potassium, Phosphorus Sensor) which aids in monitoring and identifying the deficiency of nutrients present in the soil. This sensor uses the colorimetric principle to carry out analysis on the nutrients that are present in the soil. It has light-emitting diodes and Light-dependent resistors in it. The data that is sensed by these NPK sensors are then forwarded to the cloud and fuzzification logic is applied to it as the data retrieved is in a very vague format. From the analysis carried out, an alert message is sent to the farmers to inform them the quantity of nutrients in the form of N, P, and K required by the soil at regular intervals. This helps in yielding good quality and quantity of crops. Kang, Ju-Hee et al. [18] have designed and implemented an information retrieval system that aids in retrieving information related to insect pests and diseases for the cultivation of crops and helps the users in checking real-time information with the help of their phones using Lucene which is a library that specifically works on image analysis.

3.4. Smart Transportation. With the advancement of sensors, communication technologies, automated and high-speed networking technologies, transportation and traffic management in cities have become quite modernized and smarter. Smart transportation helps in enhancing the efficiency of an individual in moving around from one place to another along with ensuring safety. The technologies facilitating this smart transportation include IoT devices and the 5G networking technology. Yongming Feng in [19] proposes a solution faced by vehicular networks in real-time information retrieval of data related to vehicular navigation and its positioning. The traditional method of information retrieval is not that efficient in fetching real-time data. The proposed solution on the analysis of the Frame difference method, the optical flow method, and the Background difference method plays a major role in the detection of moving vehicles or targets. The pros and cons of all these three methods are discussed. The author mainly focuses on the study of image features-based vehicle retrieval algorithms for proposing a solution for information retrieval of the navigation of vehicles and their positioning. Camilo Castellanos et al. [20] using different case studies represents a RA for addressing the challenges of big data analysis in STS that includes architectural patterns and different tactics.

3.5. Smart Industry. Jing Wang et al in [21] have proposed a pre-warning system for food safety that adopts association rule mining and IoT technology, for monitoring at regular intervals of time, all the detected data of the entire supply chain and automatically pre-warning the system if any abnormality is detected. K. Moorthi et al. [22] have carried out data analytics for various e-commerce companies using historic and static data. The analysis proves that data grows and changes now and then and it is a requirement of new models as well as algorithms for collecting, storing, processing, analyzing, and evaluating the data in any e-commerce related applications.

3.6. Other Case Studies. Irfan Mehmood et al. in [23] propose a solution to deal with the computational quality and storage-related challenges faced during image retrieval through smartphones in an IoT environment. The authors have proposed a deep learning-based lightweight system for energy-constrained devices. The steps mentioned in the proposed system are first detection and cropping of face regions using classifiers. Then using convolutional layers to represent faces, then indexing the big data repository, for faster comparison for real-time retrieval, and finally using Euclidean distance for finding a resemblance between the images in repositories and the concerned queries. Navjyot Kaur Walia et al. in [24] have mentioned the use of information retrieval in designing an IoT-specific application for controlling Smart Lights using Wifi. Mingliu Liu et al. in [25], after investigation of a number of IoT searching scenarios, have proposed a common model for representing recordings of sensor information. The authors propose an Indexing mechanism and a tree indexing structure for improving the efficiency and accuracy of the retrieval process and ensuring the scalability for any large-scaled data simultaneously. A large number of simulations were also carried out for demonstrating the effectiveness of their proposed solution. Ananda Ghosh et al. in [26] propose an approach based on deep learning for the reduction of data on the edge with machine learning on the cloud by investigating the merging of edge and cloud computing for IoT data analytics. To reduce the dimensions of data, the auto-encoder's encoder part is placed on the edge which helps in reducing the dimensions of data. The reduced amount of data is then transferred to the cloud from where it can be directly used for machine learning. The data can be also converted back to its original features by the auto-encoder's decoder part. The approach proposed was evaluated and the results show that the reduction of data did not have a major impact on the classification accuracy. Only a minor effect was seen on the classification accuracy even by a 77% reduction in data.

Table 3.1 summarizes the various IoT applications, data used in that application, IoT device or sensors and data analytics tool being used.

Table 3.1
IoT Data Analytics Applications

Source	IoT Application	Type of Data	Summary	IoT Device	Year	Data Analytics Tool	Pros	Cons	Future work	Methodology used
[10]	Smart City	Real Time Text	Real Time analysis of urban city data	ZigBee based Vehicular Network	2018	Apache Spark, Hadoop, Giraph	Efficient in terms of Scalability and Big Data Processing	Time complexity grows with amount of data	Integration of graph network and dynamicity	Graph Algorithms using Mapreduce
[11]	Smart City	Real Time Data	Adaptive heuristic model for traffic congestion and carbon emission	Signal processing and location based sensor	2020	Data Graph	Comprehensive Design for Smart Traffic and Congestion forecasting, Efficient Accuracy and less error rate	Societal and demographic activity factors not included like major events	Intelligent Transport Services, Driverless Technology can be mapped with IoT Devices	Probability Analysis Model
[12]	Smart City	Real Time Data	Solar power based smart bin to segregate wet or dry waste and overflow alert system	Ultrasonic Sensor, Level Sensor, weight sensor	2020	Data Graph	Dry and wet waste segregator, odour control, human detection facility, route mechanism	Solar panel battery and regular maintenance	Autonomous movable smart bin	Optimal route selection algorithm is used for route selection for bin disposal
[13]	Smart Building HVAC Fault Detection and Diagnosis	Real Time Data	Feature Extraction for HVAC fault finding and alert	Terminal Unit data of heating, cooling, deadband	2018	Apache Spark, Cassandra and MC-SVM	Effective fault diagnosis for all faulty and non faulty terminal units of HVAC	Model is application for fan coil terminal unit of single building, Non inclusion of internal and external faults	Fully remote fault diagnosis, Generalized terminal unit model can be developed	X-Means Clustering, Multi Class Support Vector Machine

Source	IoT Application	Type of Data	Summary	IoT Device	Year	Data Analytics Tool	Pros	Cons	Future work	Methodology used
[14]	Smart Building Smart Home Energy Management	Real Time Data	Energy Management for a Smart home	Home Automation Sensor	2020	J48, WEKA	Recommendation for smart home management, Demographic specific energy saving results	Application is specific to android domain only, customized energy saving options can not be added, Limited devices were considered	Model validation on many devices, Incorporation of location based devices, Incorporation of blockchain and cyber security	J48 Machine Learning algorithm for energy consumption pattern and behaviour, RuleML for energy efficient recommendation
[15]	Smart Building Thermal Comfort	Real Time Data	Optimal parameter setting for thermal comfort	Humidity and Temperature Sensor	2018	Nonlinear ML models (NN)	Better performance of non linear models compared to linear ones with minimal training time	In addition to HVAC settings, other factors impacting building energy usage should be investigated.	Other data aspects, such as occupancy and building location can be incorporated	Machine learning techniques for comfort level modelling
[16]	Smart Building Smart Control	Real Time Data	Smart control of oxygen level, smoke detection and luminosity level in building	Oxygen Sensor, Gas Detector and Luminosity sensor	2016	PySpark and Cloudera	Integration of BDA and IoT for handling the enormous volume and velocity challenge of real-time data in the smart building area	Applicable only on the Smart building domain	Applicability to other domains	Python and the Big Data Cloudera platform

Source	**IoT Application**	**Type of Data**	**Summary**	**IoT Device**	**Year**	**Data Analytics Tool**	**Pros**	**Cons**	**Future work**	**Methodology used**
[17]	Smart Agriculture Fertilizer Intimation System	Voltage based on the chemicals present in the soil	Design of a Nitrogen Phosphorus Potassium sensor.	Novel NPK Sensor	2020	Google cloud with fuzzy logic	IoT solution that is low-cost, accurate, and intelligent	The sms generated do not mention the amount or quantity of fertilizer to be added	A separate module can be added to the system mentioning the required amount of fertilizer	Fuzzy Rule based system
[18]	Retrieving information related to insect pests and diseases for cultivation of crops for u-Farm	Images of the farm	Diseases and Pest analysis is carried out and users can monitor the analysis at real time that helps in better yielding of crops	High resolution cameras	2015	Lucene library for image	Realtime information retrieval on smartphones	Could include more functionalities such as temperature humidity factor monitoring	More functionalities such as modules on irrigation can be integrated on the uFarm app	Object oriented modeling
[19]	Vehicular Navigation and Positioning	Image features based vehicle retrieval	Focus is on image features based vehicle retrieval algorithms for proposing a solution for information retrieval needed for Vehicular Navigation and Positioning	Transient phone signals, GPS Trajectory	2020	Frame Difference Method, Optical Flow method and Background difference method	Efficient real-time performance with better detection capabilities.	Absolute precision in vehicle navigation is difficult and retrieval of positional data.	The method's data bytes are very close to the actual bytes and as time passes they will essentially coincide with the actual bytes	Image Matching

Source	IoT Application	Type of Data	Summary	IoT Device	Year	Data Analytics Tool	Pros	Cons	Future work	Methodology used
[20]	Addressing Big Data Analysis in STS	Architectural Patterns and different Tactics	Represents a RA that uses different case studies for addressing BDA	RA	2021	RA for addressing the challenges of big data analysis in STS	Proposes architectures for both Big Data Analytics and Smart Transportation System	Limited capabilities over large size of data	Focus on analysis to be carried on voluminous data	Architectural patterns and tactics
[21]	Sustainable Food Supply Chain	Data for entire supply chain	Data for the supply chain is collected and pre-warning is generated if any abnormality is detected	No Specific Tool	2017	Association Rule Mining	Effective identification of safety related issues	Only limited case studies identified and worked upon	More case studies related to the supply chain can be considered	Association rule mining
[23]	Quality and storage related challenges in Image retrieval	Images from Smartphones	Solution that uses Deep learning based light weight system to solve the quality and storage related challenges during Image retrieval from smartphones	Image extraction from Smart Phones	2019	Deep Learning based light-weight system	Approach is extremely efficient both in terms of complexity and accuracy	Features still cannot be stored on small capacity devices and performance of retrieval should be robust	Analyzing image representation techniques that are based on hashing	Binary Classifier and CNN Feature extraction

4. Challenges in IoT Data Analytics. In this section we present the open issue and challenges faced in IoT based Data Analytics.

4.1. Data Acquisition and Transmission. The data generated by the IoT sensors are of different types, having different structures and different dimensions. IoT Sensor data can be structured, semi-structured, or unstructured. It can be homogeneous or heterogeneous in nature. The generated data further needs to be transmitted to the database or cloud space. To transfer such large, complex, and dynamic data requires the appropriate transmission protocol, relevant to the application on hand. For some applications, the amount of data generated varies from time to time, in the case of smart traffic, the data is generated at every fraction of a second, whereas in smart agriculture the data is generated on a daily or weekly basis. The data storage needs to be appropriately chosen which can handle the data according to application, supports scalability and heterogeneity of data. In the scenario of data analytics, data access time is also important, data storage should be able to provide fast and efficient access to the stored information.

4.2. Data Processing. Prior to performing the data analytics, the data needs to be cleaned to remove the noisy and erroneous data, remove redundant data, perform the integration of the data generated from different sources, data conversion to the form as needed for data mining, removing the data abnormalities and analysis which includes techniques like normalization, scaling. If the data is voluminous, the analysis will be delayed and might not be useful for the dynamic application. The data generated by the IoT sensors may be mapped with one another in the temporal or spatial domain, it is important to learn and know that prior to analysis. The data analysis and information retrieval can be sequential or parallel in nature, considering the amount of generated data, if it is voluminous and independent, parallel data analysis is useful. While if the data needs the time ordinance, sequential processing is useful. So it is necessary to consider the data characteristics and apply the data pre-processing techniques before performing the data analysis.

4.3. Security and Privacy. The IoT Data Analytics involves data to be transferred to either the third-party data storage or the analytics platform. The data in our IoT system can span over various categories like belonging to demographics of a person, sensitive information of a government, medical information of a patient, location of a transport vehicle, and much more. As the data moves out of the IoT ecosystem, it is a prime concern for data to maintain the integrity and be secure from unauthorized access and malicious use. This requires taking the appropriate handling mechanism like service level agreement and protocol, before sending the data. Before sending data over cloud or third party, it is necessary to check these things.

4.4. Data Analytics . The data once reached the cloud or third-party software or in the internal system, requires the correct analysis to be performed. The data analytics tool or method should be selected such that it can handle the capacity, variety, and dynamic nature of data or information. The result generated from the analysis should be properly represented and visualized for better understanding. The tool or method used for analysis must be time efficient and provide an accurate result. The tool should be robust and work in all scenarios. In the current scenario, machine learning and deep learning techniques have proliferated in the domain of data analytics. Selecting the ML or DL techniques requires an understanding of the domain. They come with their challenges of time versus accuracy tradeoff, parameter tuning, efficiency, and much more. All these things should be considered, before selecting the data analytics tool or techniques.

5. Conclusion. Data analytics and information retrieval are important for any application to infer the working of the system, to make a future decision or predict the working of the model, or to infer trends, patterns, and many more hidden features. Static data once achieved from the IoT sensor if not used for analysis purposes, remains meaningless. In this paper, the overview of how data analysis is used in various IoT applications is discussed, along with potential tools and techniques used for analysis. Further on, the issues and challenges are also discussed for future exploration.

REFERENCES

[1] Atzori, L, Antonio I, and Giacomo M., *The internet of things: A survey.*, Computer networks 54, no. 15 (2010): 2787-2805.
[2] Yaqoob, I, Ejaz A, Ibrahim A T H, Abdelmuttlib I A A, Abdullah G, M I, and Mohsen G.,*Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges.*, IEEE wireless communications 24, no. 3 (2017): 10-16.

[3] Bandyopadhyay, D, and Jaydip S,*Internet of things: Applications and challenges in technology and standardization.*, Wireless personal communications 58, no. 1 (2011): 49-69.

[4] Ou, Q, Yan Z, Xiangzhen L, Yiying Z, and Lingkang Z. ,*Application of internet of things in smart grid power transmission.*, In 2012 third FTRA international conference on mobile, ubiquitous, and intelligent computing, pp. 96-100. IEEE, 2012.

[5] Khanpara P, Lavingia K. ,*Energy conservation in multimedia big data computing and the Internet of Things—A challenge In Multimedia Big Data Computing for IoT Applications*, 2020 (pp. 37-57) Springer, Singapore.

[6] Mendez, D M., Ioannis P, and Baijian Y. ,*Internet of things: Survey on security and privacy.*, arXiv preprint arXiv:1707.01879 (2017).

[7] IoT for Sustainable Development Project,*Widgets.weforum.org*, 2021. [Online]. Available: https://widgets.weforum.org/iot4d/index.html. [Accessed: 20- Oct- 2021].

[8] Elijah, O, Tharek A R, Igbafe O, Chee Y L, and MHD N H.,*An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges.*, IEEE Internet of Things Journal 5, no. 5 (2018): 3758-3773.

[9] Chormunge S, Mehta R. ,*Comparison Analysis of Extracting Frequent Itemsets Algorithms Using MapReduce*,In Intelligent Data Communication Technologies and Internet of Things 2021 (pp. 199-210). Springer, Singapore

[10] Rathore, M. M, Anand P, Won-Hwa H, HyunCheol S, Imtiaz A, and Sharjil S. ,*Exploiting IoT and big data analytics: Defining smart digital city using real-time urban data.*, Sustainable cities and society 40 (2018): 600-610.

[11] Chang, J, Seifedine N K, and Sujatha K. ,*Review and synthesis of Big Data analytics and computing for smart sustainable cities.* , IET Intelligent Transport Systems 14, no. 11 (2020): 1363-1370.

[12] Ashwin, M., Abdulrahman S A, and Azath M. ,*IoT based intelligent route selection of wastage segregation for smart cities using solar energy.*, Sustainable Energy Technologies and Assessments 46 (2021): 101281.

[13] Dey, M, Soumya P R, and Sandra D. ,*Smart building creation in large scale HVAC environments through automated fault detection and diagnosis.*, Future Generation Computer Systems 108 (2020): 950-966.

[14] Machorro-Cano, I, Giner A-H, Mario A Paredes-V, Lisbeth R-M, José L S-C, and José O O-A. ,*HEMS-IoT: A big data and machine learning-based smart home system for energy saving.*, Energies 13, no. 5 (2020): 1097.

[15] Zhang, W, Fang L, and Rui F. ,*Improved thermal comfort modeling for smart buildings: A data analytics study*, International Journal of Electrical Power and Energy Systems 103 (2018) : 634-643

[16] Bashir, M R, and Asif Q G. ,*Towards an IoT big data analytics framework: smart buildings systems.*, In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 1325-1332. IEEE, 2016.

[17] Lavanya, G., Chellasamy R, and Pugalendhi G. ,*An automated low cost IoT based Fertilizer Intimation System for smart agriculture.*, Sustainable Computing: Informatics and Systems 28 (2020): 100300.

[18] Kang, J-H, Se-Hoon J, Sun-Sik N, Won-Ho S, and Chun-Bo S. ,*Design and implementation of produce farming field-oriented smart pest information retrieval system based on mobile for u-Farm.*, The Journal of the Korea institute of electronic communication sciences 10, no. 10 (2015): 1145-1156.

[19] Feng, Y. ,*Real Time Retrieval Technology of Vehicle Navigation and Location Information under Internet of Things Environment.*, In 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI), pp. 185-188. IEEE, 2020.

[20] Castellanos, C, Boris P, and Dario C. ,*Smart Transportation: A Reference Architecture for Big Data Analytics.*, In Smart Cities: A Data Analytics Perspective, pp. 161-179. Springer, Cham, 2021.

[21] Wang, J, and Huili Y. ,*Food safety pre-warning system based on data mining for a sustainable food supply chain.*, Food Control 73 (2017): 223-229

[22] Moorthi, K., Gaurav Dhiman, P. Arulprakash, C. Suresh, and K. Srihari. ,*A survey on impact of data analytics techniques in E-commerce.*, Materials Today: Proceedings (2021).

[23] Mehmood, I, Amin U, Khan M, Der-Jiunn D, Weizhi M, Fadi A-T, Muhammad S, and Victor H C. de Albuquerque. ,*Efficient image recognition and retrieval on IoT-assisted energy-constrained platforms from big data repositories.*, IEEE Internet of Things Journal 6, no. 6 (2019): 9246-9255.

[24] Walia, N K, Parul K, and Deepti M. ,*An IoT by information retrieval approach: Smart lights controlled using WiFi.*, In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), pp. 708-712. IEEE, 2016.

[25] Liu, M, Deshi L, Qimei C, Jixuan Z, Kaitao M, and Song Z. ,*Sensor information retrieval from Internet of Things: Representation and indexing.*, IEEE Access 6 (2018): 36509-36521

[26] Ghosh, Ananda M., and Katarina G. ,*Deep learning: Edge-cloud data analytics for IoT.*, In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1-7. IEEE, 2019.

Edited by: Katarzyna Wasielewska
Received: Jan 19, 2022
Accepted: Apr 6, 2022