

The Dynamics and Implications of the Internet of Things on Data Mining

^{1*}Bongs Lainjo & ²Hanan Tmouche

¹Cybermatic International, Canada & IBN Zohr University, Agadir, Morocco

*E-mail Corresponding Author: bsuiru@icloud.com ²tmouche@gmail.com

Article Information:

Received: May 10, 2023, Accepted: May 30, 2023, Published: June 10, 2023

Abstract

The research explores and understands the thematic dynamics of the Internet of Things (IoT) and its complementary and cross-cutting data mining (DM) platform. As part of the process, secondary data is utilized based on user-app searches generated by Google Scholar. A database is compiled, analyzed, and presented. This study also discusses the classification of data mining methods and the key data mining techniques for IoT applications. The research findings indicate that IoT continues to evolve with significant degrees of proliferation. Complementary and trailblazing data mining (DM) with more access to cloud computing platforms has accelerated the achievement of planned technological innovations. The outcome has been myriads of apps currently used in different thematic landscapes. Based on available data on user app searches, between 2016 and 2019, themes like sports, supply chain, and agriculture maintained positive trends over the four years. Moreover, the emerging Internet of Nano-Things was beneficial in many sectors. Wireless Sensor Networks (WSNs) were also emerging with more accurate and effective results in gathering information and processing data and communication technologies. However, data mining in IoT applications faces significant security, complexity, and privacy challenges. In summary, available data indicate that IoT is happening and has a significant implication for data mining. All indications suggest that it will continue to grow and increasingly affect how the world interacts with "things." A backdrop of concerns exists, from developing standard protocols to protecting individual privacy. This study recommends various potential solutions; however further studies are required to determine the practicality of the suggested solutions.

Keywords: Internet Of Things, Data Mining, Trailblazing, Ubiquitous, Wireless Sensor Networks, Evolution, Iot Applications, Cloud-Based.

1. Introduction

1.1. The Dynamics and Implications of the Internet of Things on Data Mining

The Internet of Things (IoT) can be traced back to the development of communication protocols, the creation of networks, and the development of intelligent devices. Kong et al. (2020) describe the IoT as involving the information of physical objects, including sensors, machines, cars, buildings, and other items, that allow interaction and cooperation of these objects to achieve specific goals. Others, such as Welbourne et al. (2009), provide more detailed descriptions; that is, IoT was presented as a structure of computing that facilitates data transfer between a digital and mechanical appliance without involving human interaction. The latter case makes IoT capable of data mining. For example, Hancock and Hancock (2016) mention that IoT has three critical ramifications: data collection, remote control capability, and object communication.

One important aspect of working with IoT is collecting data, but it's not enough to just have data – data sets need to be turned into meaningful information. Chen et al. (2015) conducted a literature review and found that algorithms can be used to extract useful data from IoT devices. This process, known as data mining (DM), involves analyzing large datasets and using algorithms to uncover hidden patterns and information. By exploring and transforming data, DM can reveal previously unknown trends and patterns, which can be used to make informed decisions and predictions in important situations (Lainjo, 2022).

1.2 Background

Notably, IoT and data mining complement each other. As such, developments in IoT then result in a consequent implication on data mining. The two technologies have a longstanding effect on various thematic areas, including developing IoT-enabled technologies such as sensors for data collection, IoT-based systems of communication and data transfer, and similar technologies (Wu He et al., 2014). The IoT and its impact on data mining have rich backgrounds beginning with the onset of the Internet (e.g., Xia et al., 2012; Wu He et al., 2014). Alternatively, IoT and data mining have been associated with increased security concerns and discussions over future developments, that is, future architecture, regulations and standardizations.

1.3 Settings

Jacob Morgan (as cited in Hancock & Hancock, 2016) supposed that anything that can be connected would be connected. Morgan referred to the future regarding the Internet of Things: an assertion based on an informed observation of the trend in the IoT space. For instance, Kong et al. (2020) mentioned that with the rapid development of IoT technologies, a series of national strategies such as Made-in-China 2025, American Advanced Manufacturing Partnership Program, and German Industry 4.0 have been put forward and implemented. In light of these strategies, the IoT will continue influencing significant paradigm shifts, especially in data management and applications. Notably, Liu, Nitschke, Williams, and Zowghi (2020) remark that the Internet of Things (IoT) is driving technological change and the development of new products and services that rely heavily on the data quality collected by IoT devices. These areas (see Figure 1) include IoT mining and Analytics, developing Sensors for data collection and management development, e-health, and Big Data analytics.

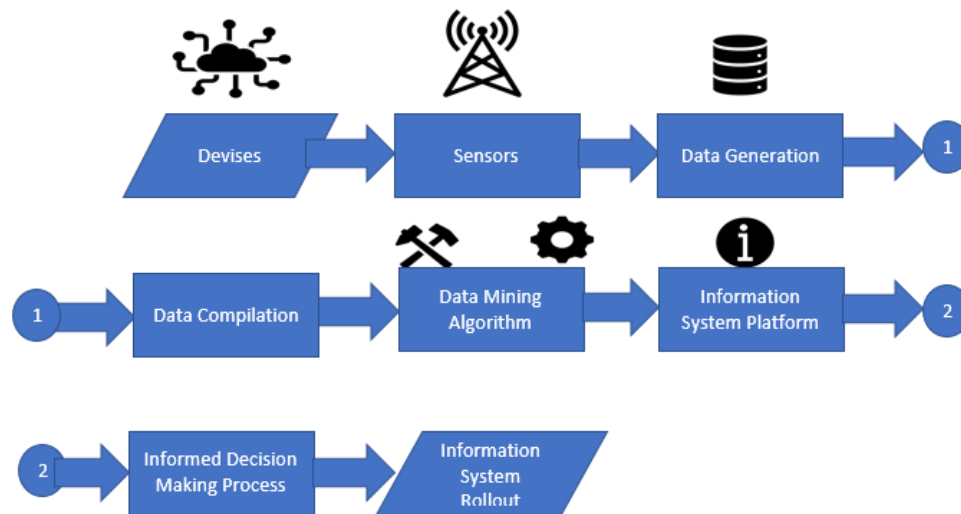


Fig 1: IoT Ecosystem Platform

Source: (Lainjo, 2022).

1.4 Prognoses

In a statement by Hancock and Hancock (2016), the authors indicate that it is hard to know the size or state of the IoT because it is evolving rapidly. A similar opinion is observed by Kong et al. (2020), where an increase in data mining and consequent demand for data storage is predicted.

2. The future of technology

The preceding literature has presented IoT and its implication for data mining as rapidly growing (Chen et al., 2015; Hancock & Hancock, 2016). Some have described it as ever-growing, while others have illustrated its future as hard to quantify due to available possibilities for growth (e.g., Hancock & Hancock, 2016). For instance, when Jacob Morgan mentions that eventually, "anything that can be connected will be connected," it signifies that the implication of IoT on data mining will be significant. For instance, even Kong et al. (2020) mentioned that with the increase in the IoT business, the demand for data storage and computing would put pressure on the ability of cloud computing.

2.1 Findings and Contribution

There are considerable implications of IoT on data mining. Some of the already experienced effects include increased efficiency in data capture and processing, more accurate statistics, and facilitation of extensive data collection and transfer, among other implications (Wu et al., 2014; Liu, Nitschke, Williams, & Zowghi, 2020). The paper also finds potential challenges, including data security and the impact on privacy through IoT-enabled data mining. This article is a contribution toward addressing current challenges and achieving some milestones.

2.2 Classification of Data Mining Techniques in IoT

Data mining methods employed in the applications of IoT are classified into two broad categories based on their execution platform. The two categories include cloud-based and onboard data mining techniques. They can also be classified based on the operation mode, i.e., streaming and batch data mining methods. Figure two depicts the two categories of the applied data mining methods in IoT.

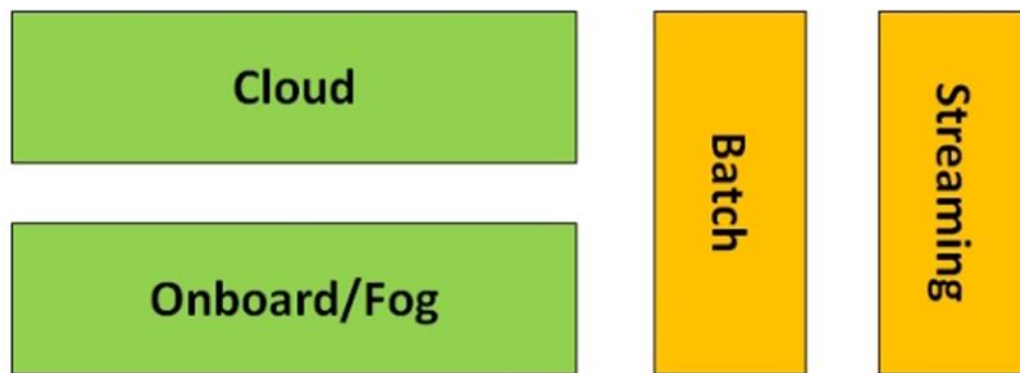
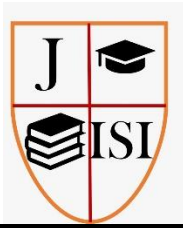


Figure 2: Classification of data mining methods in IoT

Source: (Gaber, 2012).

Onboard data mining techniques possess the power to operate in environments with limited resources. Over the past twenty years, various stakeholders have developed onboard data mining algorithms for micro-computational gadgets and wireless sensor networks (Gama & Gaber, 2007). Onboard data mining techniques are preferable for IoT applications because of the major issues associated with IoT, such as networking constraints and privacy concerns. These techniques enable IoT to become "smart objects," which are things that can detect the environment they are operating in, interpret the events taking place in the given environment, and react appropriately to them.

Cloud-based data mining techniques make large data volumes scalable by parallelizing and distributing data sets and processes. These techniques best apply to IoT applications that operate at the international level or in broad geographical regions. Information gathered and accumulated from different "things" can be applied in long-term data mining activities. Instances of such applications can include large-scale healthcare and environmental monitoring applications. Giant data mining vendors such as Google and IBM have designed and created various platforms for cloud-based data mining in IoT. Loai et al. (2016) explain how the healthcare industry uses cloud-based data mining techniques to enhance the efficiency of healthcare services. They propose that healthcare practitioners reduce latency by using *doublet* as a hardware framework between patients' mobile devices and the cloud.



The two categories discussed above arrange data mining techniques employed in IoT applications based on the location of the data mining process. Nonetheless, these techniques operate in two modes regardless of the location of the data mining process. The modes include the batch mode and the steaming mode. Batch data mining techniques operate on kept information since they are characteristically iterative (Gaber et al., 2018). These techniques are best applicable for IoT applications that operate on historical information at various magnitudes of granularity. Hence, they are characteristically cloud-based.

Steaming data mining techniques are used on active data and are best applicable in IoT applications when there is a real-time necessity to model the data mining process and high data velocity. They can be employed at the edge or cloud. Various data scientists and data analysts have proposed multiple data streaming techniques over the last twenty years. M.M. Gaber (2012) categorizes the techniques into four distinct categories: Granularity-based methods, symbolic approximation-based techniques, Hoefding bound-based techniques, and two-phase methods. Hoefding bound-based methods use statistical measures to establish a sample size employed in various ways depending on the adopted method. On the other hand, granularity-based methods make appropriate adjustments depending on the computational resources available (Shobanadevi & Maragatham, 2017). Symbolic approximation-based techniques convert time series into dense symbolic rendition, while two-phase techniques use an online processing stage to feed in a batch stage.

However, synthesizing or selecting the most practical data mining algorithm remains a challenging feat for all smart environments with IoT. The practicality of the algorithm is based on its capacity to create valuable analytics, precisely forecast future occasions, and handle the services and network effectively without violating the set constraints (Gupta & Chandra, 2020). Figure 3 shows the numerous IoT applications on both large and small scales. The blocks represent various IoT applications that carry out specific tasks in different fields or sectors.

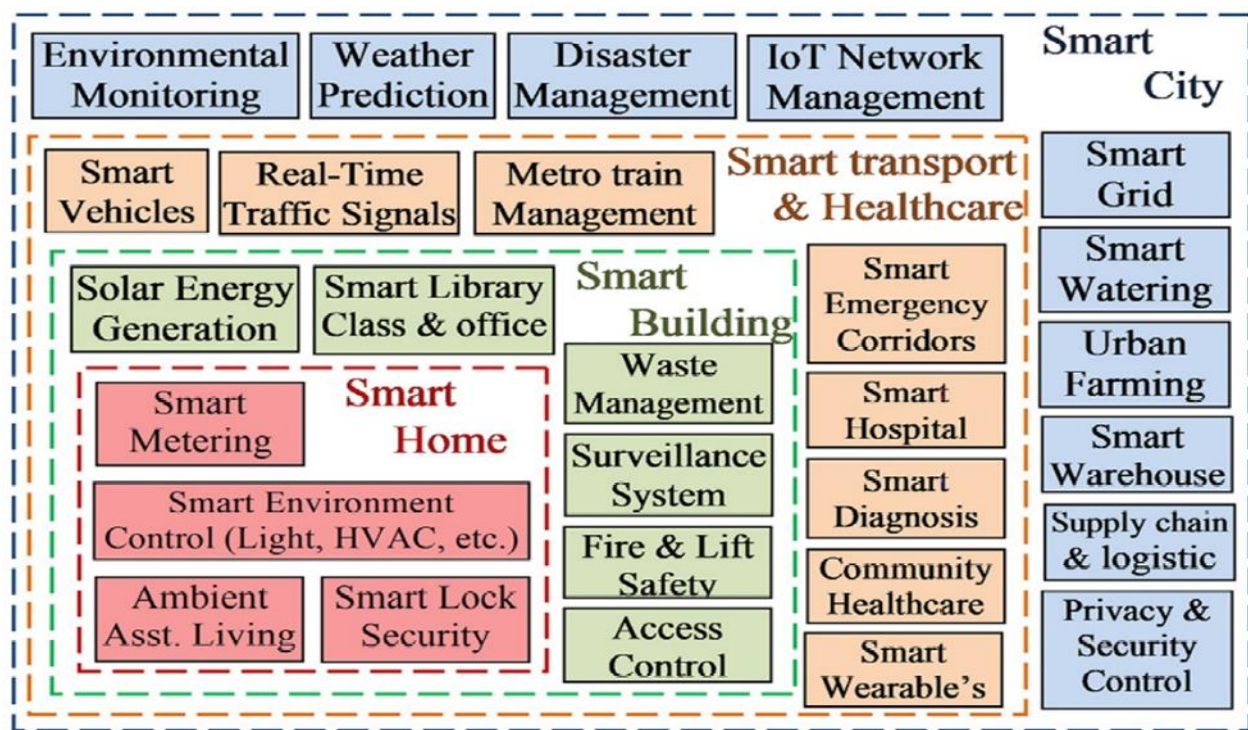


Fig 3: Large-scale and small-scale IoT applications , Source: (Gaber, 2012).

2.3 Key Data Mining Techniques Used in Big Data IoT Applications

Figure 4 shows that information gathered from different IoT gadgets is initially transited to a pre-processing section where they undergo multiple processes, such as noise abstraction, to design the basic information into the desired format. The formatted data is then transited to the data mining section, where data analysts use different methods to extract useful insights from the formatted data (Plotnivoka et al., 2020). Data mining and pre-processing sections are classified under a single unit called DL. The product of DL is examined and interpreted into knowledge that can be further interpreted by machines and understood by human beings. The knowledge is further utilized by the IoT framework, as depicted below.

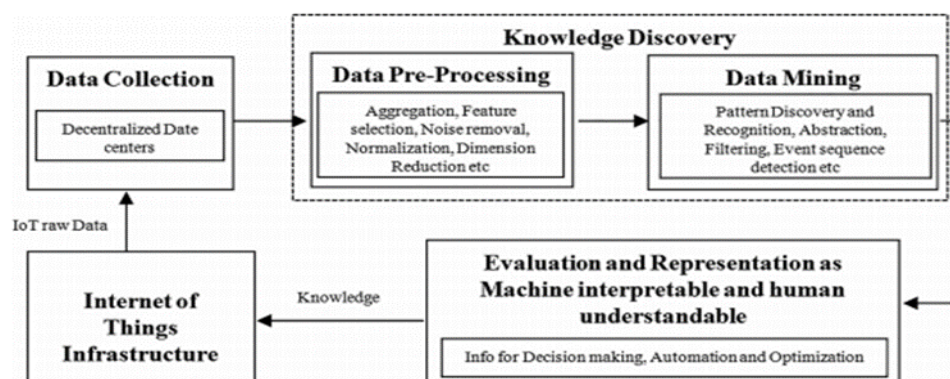


Fig 4: Overview of knowledge discovery , Source: (Sunhare et al., 2020).

Data mining begins immediately after the knowledge discovery phase. The first data mining technique is classification, which allocates data objects to classes defined before the classification process begins. Its main objective is to forecast the destination class for all data objects. The second technique is clustering. A cluster contains a team of similar data objects. Clustering employs algorithms that categorize the gathered data objects into a specific volume of groups where the data objects in a given group possess identical aspects (Ageed et al., 2021). Other key data mining techniques in IoT applications include frequent pattern, association analysis, and deviation-based outlier detection. Figure 6 shows the key data mining methods used in IoT and how they interact or relate.

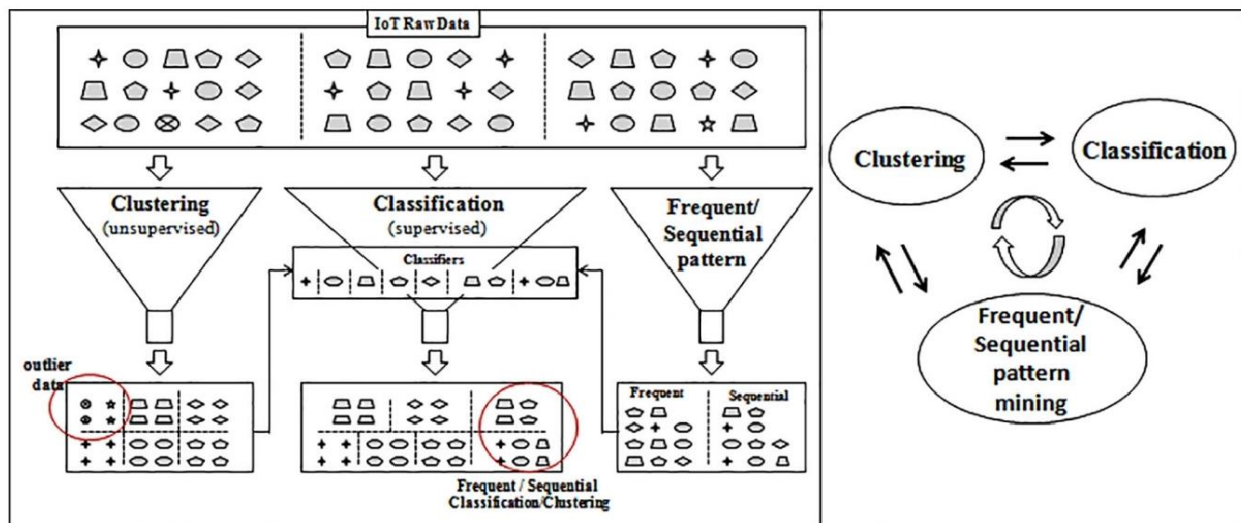


Figure 5: Data mining techniques for IoT applications.

Source: (Sunhare et al., 2020).

2.4 Challenges Facing Data Mining in IoT

In recent decades, the market for IoT applications has grown exponentially, but various stakeholders, such as manufacturers, consumers, and data scientists, have experienced significant setbacks concerning data mining in IoT applications. One of the most significant challenges is data security. Some IoT applications gather substantially sensitive data. For instance, IoT devices gather protected health information (PHI) from patients and healthcare practitioners in the healthcare sector. IoT applications such as voice assistants, internet-enabled cameras, and motion sensors can monitor and gather information about people's conversations and activities (Sunhare et al., 2020). In the Manufacturing industry, IoT applications have access to highly sensitive data regarding the production procedures and manufacturing processes. Ensuring the security of this information has become a common setback for IoT applications. Most of these applications are designed to be easily accessed by relevant stakeholders from the public internet as they are required to transfer data to cloud-based servers for data mining. Their uses also require them to be accessible

from web-based portals and mobile devices. This demand severely lowers their security (Javid et al., 2020). Cybercriminals take advantage of these security issues and gain unauthorized access to the sensitive data in the devices. The security of the data lowers as it moves from its source, such as sensors, to the cloud-based servers for data mining. Figure three shows the security levels of IoT applications.

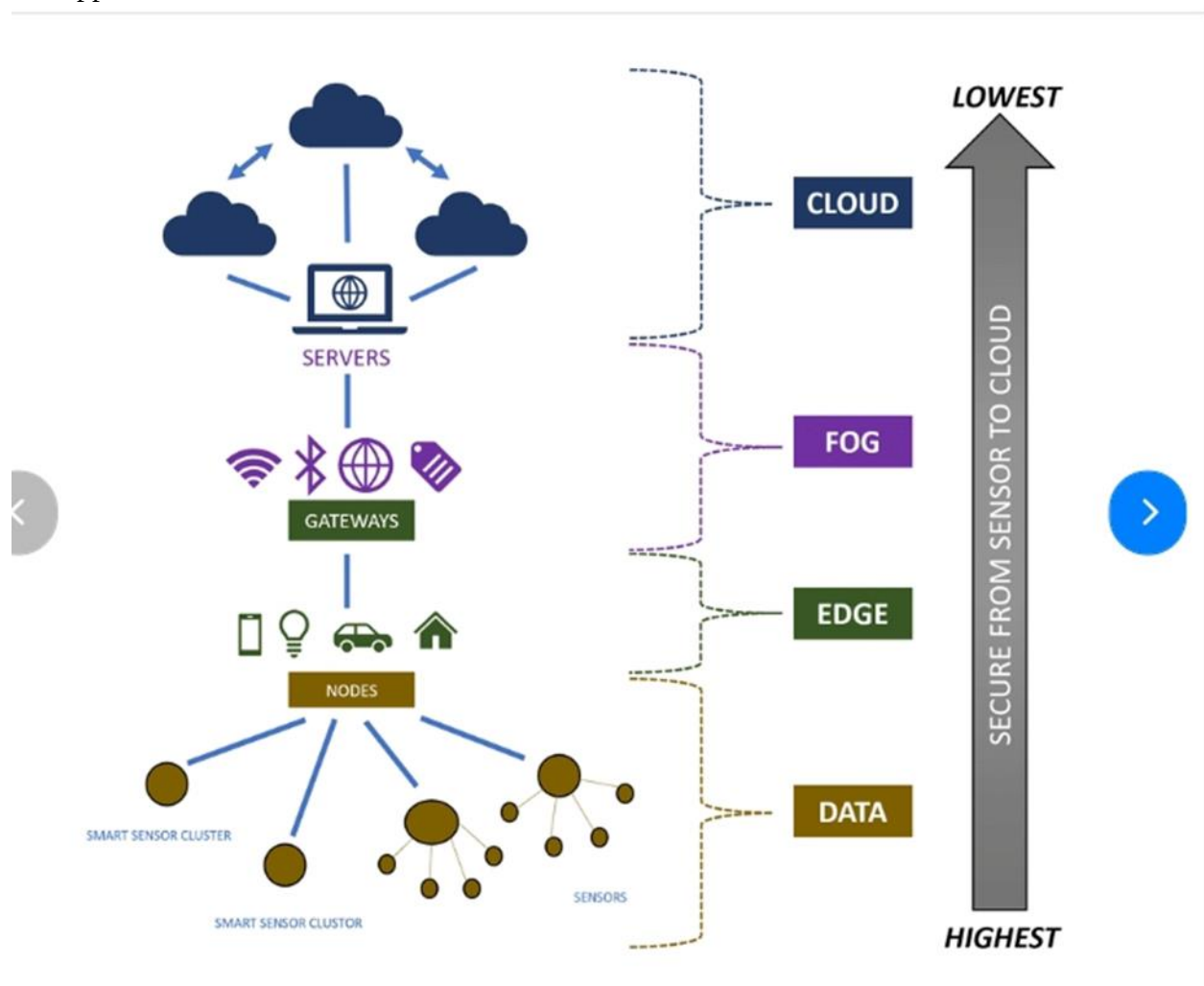


Figure 6: Security levels of IoT applications

Source: (Gunathilake, 2019).

Data privacy is another significant setback facing data mining in IoT applications. The data gathered by IoT devices and sent to cloud-based servers for data mining are generally protected by different information privacy legislations. Sometimes IoT applications unintentionally violate the regulations as they may gather personal information without users' consent. For instance, voice assistants can sometimes overhear personal conversations and store them without the user's consent. Additionally, information protection laws require data encryption during data storage and transition. IoT applications often have inadequate data mining resources and power; hence, they

experience difficulty encrypting data. Numerous data protection laws are strict about jurisdiction (Osamy & Khedr, 2022). For instance, the General Data Protection Regulation (GDPR) of the European Union prohibits the transition of EU citizens' data to nations that have not implemented appropriate information protection regulations. However, IoT applications need more power to track and limit data flows successfully.

The third challenge is large data volumes. IoT applications increase daily and gather massive volumes of information every second. In 2019, IoT applications created approximately 18.3 zettabytes of information, and studies forecast that the volume might reach 74 zettabytes by 2025 (Sestino et al., 2020). Carrying out data mining processes such as transmitting, storing, and processing vast volumes of data are complex, difficult, and expensive. This is because such vast volumes of data require cloud-based servers that can rapidly perform data mining processes while sending the necessary commands or alerts to the IoT applications.

The fourth setback facing data mining in IoT is complexity of data. Multiple IoT applications are designed to operate based on a Big Data mentality. They gather as much data as possible and transit it to cloud-based data mining servers. This approach leads to the production of vast volumes of data and generates complex datasets. The information gathered from IoT devices needs more insight and is mostly unstructured. As a result, data analysts and data scientists are forced to timestamp, index, and correlate the data with other data sources to achieve the desired outcomes. In combination with data complexity, vast data volumes make it challenging to carry out data mining efficiently and effectively in IoT applications. Multiple data mining techniques and tools designed to handle complex datasets cannot keep up with the vast volumes of information produced by IoT devices (Ibrahim, 2018). Contrarily, the data mining techniques and tools that can manage Big Data often fail to provide the desired degree of in-depth analysis or may fail to meet the latency demands of IoT applications.

2.5 Potential Solutions to the Challenges Facing Data Mining in IoT Applications

IoT applications gather and process vast volumes of complex information; hence they must be protected against cyber-attacks and secured under information privacy legislation. The challenges discussed above are significant, but there are potential solutions. Relevant stakeholders can take advantage of the new generation 5G mobile networks, which offer IoT applications the operational capacity and the bandwidth to transfer and process vast volumes of data (Yang, 2022). 5G mobile networks also offer scalable cloud infrastructure that meets the demands for processing complex data while guaranteeing data privacy and security.

2.6 Contribution to Future Research

This paper's recommendations concerning the potential solutions to data mining challenges in IoT applications are based on a limited number of studies (Kumar et al., 2019). Therefore, future studies should build on these potential solutions to establish whether they are practical. Additionally, this paper has discussed only two potential solutions to the challenges, but since data mining and IoT applications are dynamic fields, future studies should find more solutions to these setbacks.

2.7 Limitation

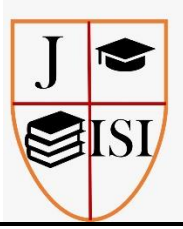
This paper employs qualitative research methods, implying that the information used for this research was obtained from secondary sources. Its most significant limitation is that there are limited peer-reviewed scholarly articles regarding the topic of study. Therefore, more research concerning the dynamics and implications of the Internet of Things on data mining is needed to approve or disapprove the findings of this paper.

3. Conclusion

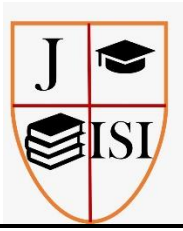
In summary, the preceding discussion has indicated that IoT is happening and has a significant implication for data mining. All indications suggest that it will continue to grow and increasingly affect how we interact with "things." A backdrop of concerns exists. Some of the matters raised by researchers include the technology's growth rate against the ability to regulate implementation, challenges in standardization, information integrity, and data security (He et al., 2018; Tsai et al., 2014). Accordingly, the implications of IoT on data mining will continue to be felt in what can be described as unprecedented. As mentioned in the paper, as many "things" continue to be connected and communicate through IoT, a significant effect will continue to be noticed in data mining.

References

- Agreed, Z. S., Zeebaree, S. R., Sadeeq, M. M., Kak, S. F., Yahia, H. S., Mahmood, M. R., & Ibrahim, I. M. (2021). A comprehensive survey of big data mining approaches in cloud systems. *Qubahan Academic Journal*, 1(2), 29-38.
- Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks*, 11(8), 431047. <https://doi.org/10.1155/2015/431047>
- Data mining: concepts, models, and techniques. (2011), 49(04), 49-2107-49-2107. <https://doi.org/10.5860/choice.49-2107>
- Gaber, M. M. (2012). Advances in data stream mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2 (1), 79–85.
- Gaber, M. M., Aneiba, A., Basurra, S., Batty, O., Elmisery, A. M., Kovalchuk, Y., & Rehman, M. H. U. (2018). Internet of Things and data mining: From applications to techniques and systems. *WIREs Data Mining and Knowledge Discovery*, 9(3). <https://doi.org/10.1002/widm.1292>
- Gama, J., & Gaber, M. M. (2007). Learning from data streams: processing techniques in
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660. <https://doi.org/10.1016/j.future.2013.01.010>
- Gunathilake, N. (2019). Security levels of IoT architecture [Online Image]. In *Research Gate*. https://www.researchgate.net/figure/Security-levels-of-IoT-architecture_fig1_331298498



- Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, 12(4), 1243-1257.
- Hancock, B., & Hancock, L. (2016). Somebody's Watching: The Ever-Growing Internet of Things. *Phi Kappa Phi Forum*, 96(3), 12–15.
- Hand, D. (2007). Principles of Data Mining. *Drug Safety*, 30(7), 621-622. <https://doi.org/10.2165/00002018-200730070-00010>
- He, Y., Guo, J., & Zheng, X. (2018). From Surveillance to Digital Twin: Challenges and Recent Advances of Signal Processing for Industrial Internet of Things. *IEEE Signal Processing Magazine*, 35(5), 120-129. <https://doi.org/10.1109/msp.2018.2842228>
- Ibrahim, N. (2018). Publication and Discovery of Things in the Internet of Things. *EAI Endorsed Transactions on Internet of Things*, 4(14), 156082. <https://doi.org/10.4108/eai.20-12-2018.156082>
- Javid, T., Gupta, M. K., & Gupta, A. (2020). A hybrid-security model for privacy-enhanced distributed data mining. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.06.010>
- Kong, D., Liu, D., Zhang, L., He, L., Shi, Q., & Ma, X. (2020). Sensor anomaly detection in the industrial Internet of Things based on edge computing. *Turkish Journal of Electrical Engineering & Computer Sciences*, 28(1), 331–346. doi:10.3906/elk-1906-55
- Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0268-2>
- Lainjo, B. (2022). The Dynamics of the Ubiquitous Internet of Things (IoT) and Trailblazing Data Mining (DM). *International Journal of Wireless & Mobile Networks*, 14(3). <https://doi.org/10.5121/ijwmn.2022.14302>
- Loai, A. T., Mehmood, R., Benkhelifa, E., & Song, H. (2016). Mobile cloud computing model and big data analysis for healthcare applications. *IEEE Access*, 4, 6171–6180
- Osamy, W., & Khedr, A. M. (2022). FACS: Fairness-aware clustering scheme for monitoring applications of Internet of things based wireless sensor networks. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part B), 3615–3629. <https://doi.org/10.1016/j.jksuci.2022.03.030>
- Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: a systematic literature review. *PeerJ Computer Science*, 6, e267. sensor networks. Springer
- Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. *Technovation*, 98, 102173. <https://doi.org/10.1016/j.technovation.2020.102173>
- Shobanadevi, A., & Maragatham, G. (2017, December 1). *Data mining techniques for IoT and big data — A survey*. IEEE Xplore. <https://doi.org/10.1109/ISS1.2017.8389260>



- Sunhare, P., Chowdhary, R. R., & Chattopadhyay, M. K. (2020). Internet of things and data mining: An application-oriented survey. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.07.002>
- Tsai, C., Lai, C., & Vasilakos, A. (2014). Future Internet of Things: open issues and challenges. *Wireless Networks*, 20(8), 2201-2217. <https://doi.org/10.1007/s11276-014-0731-0>
- Welbourne, E., Battle, L., Cole, G., Gould, K., Rector, K., & Raymer, S. et al. (2009). Building the Internet of Things Using RFID: The RFID Ecosystem Experience. *IEEE Internet Computing*, 13(3), 48-55. <https://doi.org/10.1109/mic.2009.52>
- Wu He, Gongjun Yan, & Li Da Xu. (2014). Developing Vehicular Data Cloud Services in the IoT Environment. *IEEE Transactions on Industrial Informatics*, 10(2), 1587-1595. <https://doi.org/10.1109/tii.2014.2299233>
- Wu, Q., Ding, G., Xu, Y., Feng, S., Du, Z., Wang, J., & Long, K. (2014). Cognitive Internet of Things: A New Paradigm Beyond Connection. *IEEE Internet Of Things Journal*, 1(2), 129-143. <https://doi.org/10.1109/jiot.2014.2311513>
- Xia, F., Yang, L., Wang, L., & Vinel, A. (2012). Internet of Things. *International Journal Of Communication Systems*, 25(9), 1101-1102. <https://doi.org/10.1002/dac.2417>
- Yang, G. (2022). An Overview of Current Solutions for Privacy in the Internet of Things. *Frontiers in Artificial Intelligence*, 5. <https://doi.org/10.3389/frai.2022.812732>
- Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for Smart Cities. *IEEE Internet Of Things Journal*, 1(1), 22-32. <https://doi.org/10.1109/jiot.2014.2306328>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4>)