



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Application of Data Mining Techniques in IoT

Gayatri Divekar<sup>1</sup>, Anushree Yede<sup>2</sup>, Prof. Mohit K. Popat<sup>3</sup>

B.E. Student, Department of Computer Science & Engineering, JDIET Yavatmal, Maharashtra, India<sup>1,2</sup>

Associate Professor, Department of Computer Science & Engineering, JDIET Yavatmal, Maharashtra, India<sup>3</sup>

**ABSTRACT:** Data mining isn't a new invention that came with the digital age. The concept has been around for over a century, but came into greater public focus in the 1930s. One of the first instances of data mining occurred in 1936, when Alan Turing introduced the idea of a universal machine that could perform computations similar to those of modern-day computers. Data mining is nothing but extracting knowledge from large amount data. This branch of data science derives its name from the similarities between searching for valuable information in a large database and mining a mountain for ore. Both processes require sifting through tremendous amounts of material to find hidden value. Data mining offers many applications in business. For example, the establishment of proper data (mining) processes can help a company to decrease its costs, increase revenues, or derive insights from the behaviour and practices of its customers. Certainly, it plays a vital role in the business decision-making process nowadays. Data mining is also actively utilized in finance.

## I. INTRODUCTION

Data mining is the process of uncovering patterns and finding anomalies and relationships in large datasets that can be used to make predictions about future trends. The main purpose of data mining is extracting valuable information from available data. Data mining is considered an interdisciplinary field that joins the techniques of computer science and statistics. Note that the term "data mining" is a misnomer. It is primarily concerned with discovering patterns and anomalies within datasets, but it is not related to the extraction of the data itself. The field is rapidly evolving. New data emerges at enormously fast speeds while technological advancements allow for more efficient ways to solve existing problems. In addition, developments in the areas of artificial intelligence and machine learning provide new paths to precision and efficiency in the field. There is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyse this huge amount of data and extract useful information from it. Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users. The data mining process breaks down into five steps. First, organizations collect data and load it into their data warehouses. Next, they store and manage the data, either on in-house servers or the cloud. Business analysts, management teams and information technology professionals access the data and determine how they want to organize it. Then, application software sorts the data based on the user's results, and finally, the end-user presents the data in an easy-to-share format, such as a graph or table. The fields include: education, healthcare, web, data stream and big data. We begin with an overview of research methodology, basic data mining processes and methods. Then we divided the selected research fields into three groups according to the application type, data source or data type. In each research area, we present the processes, algorithms, applications, and challenges from the relevant surveys and research. At the end of this paper, we discuss the conclusion and future directions based on the completed research[12].

## II. RELATED WORK

IoT has emerged as the most interactive and social area currently attracting the youth. IoT comprises of objects or things that are easily accessible and are using internet, RFID as communication means [1]. RFID technology was used earlier and required a tag and a reader to fetch the information [2]-[3]. In [4], it is mentioned that RFID's application in IoT is in the area of inventory or supply chain management. Sensor technology is latest trend in sector of IoT with almost all applications using it. These technologies like Bluetooth, RFIS, sensor helps in first phase of data collection, the next phase is to extract the meaningful information by applying DM. Although DM is implicative in IoT

applications in same way as it is used in other areas, but still it comes with certain challenges with IoT due to below stated reasons [5]-[6]:

1) Huge Data Collection by Distributed Devices: With number of devices used for IoT operation, each device generates bulk of data resulting in continuous data of huge volume.

2) No Assured Security: Different start-up companies offering IoT solutions may not focus on security. This will result in collection of huge data even from the un-trusted sources wasting the time and hence lowering the efficiency of DM technique.

3) Data Heterogeneity: Distributed IoT devices collect information from different sources of varying formats resulting in data heterogeneity making the task of DM challenging.

4) Constrained Devices: Expecting high performance from the constrained devices in terms of power or capacity is the need o IoT. This refrains IoT from having a standard integration of security leading to generation of illegitimate data.

5) Real Time Operation: IoT devices collect information at a specific time slot and tend to work in real time. Therefore, time stamps need to be recorded and decisions should be generated using DM technique in real time making it a challenge for DM technique.

6) Noise Information: IoT application collect tiny information that result in noise or error in while collection or transmission. Considering all these challenges, this research paper presents different DM techniques available for IoT applications. Firstly, the different DM techniques available in literature are discussed. Next, advantages and disadvantages of using a DM technique in applications of IoT are reflected. In last section, a performance analysis is made on basis of precision, recall, and accuracy [7].

### III. VARIANTS OF DATA MINING MODEL FOR THE IOT

#### A. Multi-Layer Data Mining Model:

As shown in Figure 1, model is partitioned into four layers namely information gathering layer, information administration layer, event processing layer and information mining service layer. Among them, information accumulation layer embraces devices , e.g. RFID sinks/readers and so forth., to gather intelligent information from different objects such as RFID stream information, GPS information, satellite information, positional information and sensor information and so on. Extraordinary sort of information requires distinctive information storage methodology. In the procedure of information collection, a progression of issues, e.g., vitality output, fault tolerance, data preprocessing, communications and so forth., ought to be very much explained [8]Information i.e. data management layer focuses in centralized or disseminated database or information stockroom i.e. data warehouse to oversee gathered information. An Event is a combination that joins information, time and different variables, so it gives an abnormal state component to information handling of IoT. An Event handling layer is utilized to investigate an events in IoT viably. Hence, it enables querying or investigation based on an event at this layer [9].

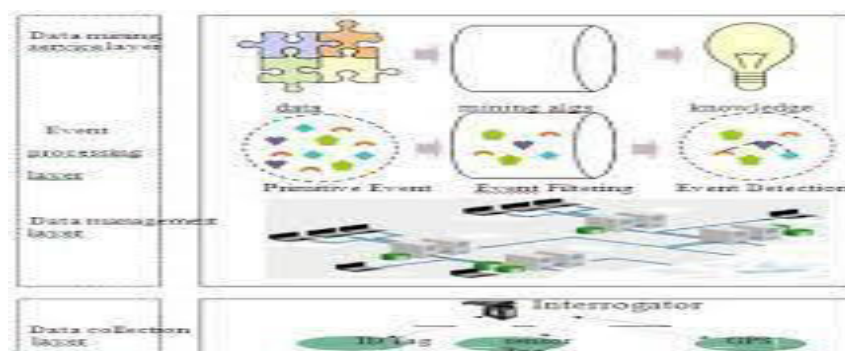
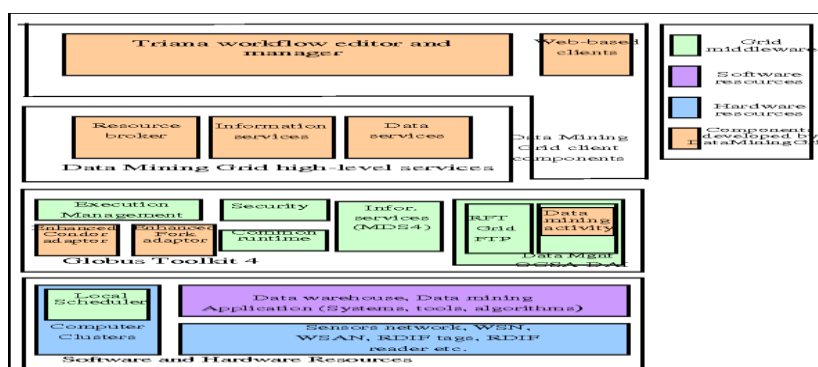


Figure 1. Multi Layer Model [10]

**B. Distributed data mining model :**

Comparing to traditional information which is raw in nature, Information in IoT has its own attributes. For instance, the information in IoT is dependably mass, appropriated, timerelated and position-related. At the same time, the information wellsprings of IoT are heterogeneous, and the assets of nodes are restricted. These attributes bring a few issues to unified information mining design. At initially, mass information of IoT is put away in distinctive locales. Consequently, it is troublesome for us to mine conveyed information by concentrated engineering. Furthermore, information in IoT is mass and needs preprocessing continuously. For the thought of information security, information protection, adaptation to internal failure, business rivalry, legitimate requirements and different elements, the technique of assembling every pertinent datum is regularly not doable. In addition, the assets of nodes are restricted. The technique of transferring all information to central nodes does not enhance the utilization of vitality expensive transmissions. In most cases, the central node needn't bother with all information, yet a few evaluations of parameters. Hence, we can preprocess the raw information in the appropriated distributed node, and afterward send the fundamental information to the recipient. Appropriated information digging model for IoT isn't just capable to take care of the issues brought by distributed capacity of nodes, but also decomposes problem complexity. In this way, the necessity of elite performance, high storage computing and processing power is decreased. In this model, the global control node is the center of the entire information mining framework. It picks the information mining calculation and the informational collections for mining, and afterward explores to the sub-nodes containing these informational collections. The sub-nodes get the raw information from different savvy objects. These raw information is supplied as an input to data filtration for preprocessing and then data abstraction and data compression, and finally, it get stored in the local data warehouse. Event separation, recognition and data mining at local nodes results into local models. Global models are the result of aggregation of local models are aggregated Subnodess trade protest information, process information and learning with each other. The entire procedure is restricted by the multi-agent based collaborative management module which is depicted in figure 2[11]

**Figure 2. Distributed data mining model [10]****IV. DATA MINING APPLICATIONS**

We divided the five research fields into three groups according to the application type, data source, or data type. This section presents educational data mining and healthcare data mining as the two most active applications in data mining fields. The remaining research fields are represented in the next two sections, where the web data is considered as one of the data sources and big data and data stream are considered as two of the data types.

**A. Educational Data Mining:**

Educational Data Mining (EDM) is the process of extracting knowledge from large educational data sets. Data from the educational environment may originate from students, instructors or management feedback or observations, and can be utilized to address important educational questions. There are several studies on EDM applications, algorithms and latest trends [12] [13] [14] [15] [16] [17]. Also[18] and [19] are related to the visualization in educational data mining. In [20], authors propose a unified data framework to aggregate data from



several sources related to higher education into one clean, amalgamated database. 37 There are many popular methods within the educational data mining field. These methods can be divided into two types of general methods: ones that can be used in EDM and other data mining fields such as prediction, clustering, outlier detection, relationship mining, social network analysis, process mining and text mining. Others are specific to the EDM field like distillation of data for human judgment, discovery with models, knowledge tracing and nonnegative matrix factorization [21][12].

#### B. Data Mining in Healthcare:

Data mining in healthcare has several goals, such as diseases prediction and diagnoses, enhancing medical services, medical costs lowering, identifying patients' risk and several other goals. Patients, hospitals and insurance companies are the main beneficiaries of the mining results. Data resources include medical reports, management reports, and any related information. There are several studies on data mining applications and algorithms in the healthcare field. Shivakumar and Alby [22] provided a survey of data mining methods applied in the previous applications related to diabetes prediction and diagnosis. Sumalatha and Muniraj [23] presented and analysed several previous works in three diseases: heart disease, breast cancer disease, and diabetes disease. Palivela et al. [24] proposed and evaluated a hybrid model that helps doctors in breast cancer diagnosis and treatment planning. Herland et al. [25] surveyed recent studies on big data in health informatics. The most commonly used methods in the previous data mining applications in the healthcare field [26] are classification, regression, clustering and association. On the other hand, the most common algorithms are K-Nearest Neighbour (K-NN), Decision Tree (DT), Support Vector Machine (SVM), Neural Network (NN), linear regression, non-linear regression, partitioned clustering, Apriori and Frequent Pattern Tree algorithm.[27]

#### C. Data Mining in Industry:

Data mining can highly benefit industries such as retail, banking, and telecommunications; classification and clustering can be applied to this area [28]. One of the key success factors of insurance organizations and banks is the assessment of borrowers' credit worthiness in advance during the credit evaluation process. Credit scoring becomes more and more important and several data mining methods are applied for credit scoring problem [29–30]. Retailers collect customer information, related transactions information, and product information to significantly improve accuracy of product demand forecasting, assortment optimization, product recommendation, and ranking across retailers and manufacturers [31, 32]. Researchers leverage SVM [33], support vector regression [34], or Bass model [35] to forecast the products' demand [36]

### V. KEY ISSUES IN DATA MINING OF IOT

There are various issues involved in data mining in Internet of Things:

#### A. Efficiency in data gathering:

Energy efficiency, scalability and fault tolerance should be taken into consideration when data is to be collected from distributed sensor networks [37] .

#### B. Data abstraction and aggregation:

Managing massive data generated from IoT is a challenging task. Efficient mechanism should be adopted for data deduplication.

#### C. Distributed data processing and mining:

Due to nodes' constraints, paradigm shift is needed for prior level preprocessing of the data at each distributed nodes and aggregated information is to be sent to sink node in order to optimize energy usage instead of sending all distributed data to server for processing.

#### D. Data mining towards the next age of Internet:

In an upcoming generations of Internet, latest trends and technologies like ubiquitous computing, semantic web, IPv6 technologies are going to be integrated with IoT. This will give rise to challenges for Data Mining due to heterogeneous unstructured data [38][11].

## VI. CONCLUSION

In the future, **data mining** will include more complex **data** types. In addition, for any model that has been designed, further refinement is possible by examining other variables and their relationships. Research in **data mining** will result in new methods to determine the most interesting characteristics in the **data**. The IoT has the potential to dramatically increase the availability of information, and is likely to transform companies and organizations in virtually every industry around the world. We all know there are huge benefits associated with IoT in different sectors. To reap the benefits from IoT, the company needs to have proper portal where the information is gathered and quickly work on those data. Companies that react faster to the data saw good returns on investment which is very important for every industry. So we use data mining concept to make IoT more smarter.

## REFERENCES

- [1] European Commission, Internet of Things in 2020: A roadmap for the future. 2008.
- [2] R. Parks and T. Pennsylvania, "RFID Privacy Issues in Healthcare: Exploring the Roles of Technologies and Regulations," pp. 1–24
- [3] F. Mattern and C. Floerkemeier, "From the internet of computers to the internet of things," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6462 LNCS, pp. 242–259, 2010.
- [4] L. Muguira, J. Ruiz-de-Garibay, and J. I. Vazquez, "Interacting with the digital world through RFID-powered gadgets," SoftCOM 2009 - 17th Int. Conf. Software, Telecommun. Comput. Networks, 2009.
- [5] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong, "Data mining for the internet of things: literature review and challenges." International Journal of Distributed Sensor Networks 11, no. 8 (2015): 431047.
- [6] M. Conti, S. K. Das, C. Bisdikian, M. Kumar, L. M. Ni, A. Passarella, G. Roussos, G. Tröster, G. Tsudik, and F. Zambonelli, "Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber-physical convergence." Pervasive and Mobile Computing 8, no. 1 (2012): 2-21.
- [7] Isha, Dr. Sahil Verma, kavita, Performance Analysis of Data Mining Techniques in IoT, 4th International Conference on Computing Sciences
- [8] Sebastian Scholze Claudio Cenedese Oliviu Matei, Carmen Anton. Multi-layered data mining architecture in the context of internet of things. In IEEE. IEEE, 2017.
- [9] Aashi Singh and Shilpi Sharma. Analysis on data mining models for internet of things. In I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(ISMAL), 2017 International Conference on, pages 94–100. IEEE, 2017.
- [10] Shen Bin, Liu Yuan, and Wang Xiaoyi. Research on data mining models for the internet of things. In Image Analysis and Signal Processing (IASP), 2010 International Conference on, pages 127–132. IEEE, 2010
- [11] Akshat Savaliya<sup>\*1</sup>, Aakash Bhatia<sup>2</sup>, Jitendra Bhatia<sup>3</sup>, Application of Data Mining Techniques in IoT: A Short Review, National Conference on Advanced Research Trends in Information and Computing Technologies (NCARTICT-2018), Department of IT, L. D. College of Engineering, Ahmedabad, Gujarat, India In association with International Journal of Scientific Research in Science, Engineering and Technology
- [12] C. Romero and S. Ventura, "Data mining in education," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 3, no. 1, pp. 12–27, 2013.
- [13] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," Procedia-Soc. Behav. Sci., vol. 97, pp. 320–324, 2013.
- [14] R. A. Huebner, "A survey of educational data-mining research," Res High Educ J, vol. 19, pp. 1–13, 2013.
- [15] O. G. McGrath, "Mixing and matching usage data: techniques for mining varied activity data sources," in Proceedings of the 41st annual ACM SIGUCCS conference on User services, 2013, pp. 119–122.
- [16] A. Pena-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," Expert Syst. Appl., vol. 41, no. 4, pp. 1432–1462, 2014.
- [17] A. Pena-Ayala and L. Cardenas, "How Educational Data Mining Empowers State Policies to Reform Education: The Mexican Case Study," in Educational Data Mining, A. Pena-Ayala, Ed. Springer International Publishing, 2014, pp. 65–101.
- [18] S. Gama and D. Gonçalves, "Visualizing Large Quantities of Educational Data Mining Information," in 18th International Conference on Information Visualisation (IV), 2014, pp. 102–107.
- [19] C. Antunes, "Acquiring Background Knowledge for Intelligent Tutoring Systems.," in EDM, 2009, pp. 18–27.
- [20] L. B. Ngo, V. Dantuluri, M. Stealey, S. Ahalt, and A. Apon, "An architecture for mining and visualization of us higher educational data," in Ninth International Conference on Information Technology: New Generations (TTNG), 2012, pp. 783–789.



- [21] [IS] N. Bousbia and 1. Belamri, "Which Contribution Does EDM Provide to Computer-Based Learning Environments?," in Educational Data Mining, A. Pena-Ayala, Ed. Springer International Publishing, 2014, pp. 3-28. 6
- [22] B. L. Shivakumar and S. Alby, "A Survey on Data-Mining Technologies for Prediction and Diagnosis of Diabetes," in International Conference on Intelligent Computing Applications (ICICA), 2014, pp. 167-173.
- [23] G. Sumalatha and N. J. R. Muniraj, "Survey on medical diagnosis using data mining techniques," in International Conference on Optical Imaging Sensor and Security (ICOSS), 2013, pp. 1-8.
- [24] H. Palivela, H. K. Yogish, S. Vijaykumar, and K. Patil, "Survey on mining techniques for breast cancer related data," in International Conference on Information Communication and Embedded Systems (ICICES), 2013, pp. 540-546.
- [25] M. Herland, T. M. Khoshgofiaar, and R. Wald, "Survey of Clinical Data Mining Applications on Big Data in Health Informatics," in 12th International Conference on Machine Learning and Applications (ICMLA), 2013, vol. 2, pp. 465-472.
- [26] D. Tomar and S. Agarwal, "A survey on Data Mining approaches for Healthcare," Int. J. Bio-Sci. Bio-Technol., vol. 5, no. 5, pp. 241- 266, 2013. [21] D. Sridevi, D. A. Pandurangan, and D. S. Gunasekaran, "Survey on
- [27] Ameera M. Almasoud1 , 2 , Hend S. Al-Khalifa1 and Abdulmalik Al-Salman2, Recent Developments in Data Mining Applications and Techniques, The Tenth International Conference on Digital Information Management (ICDIM 2015).
- [28] N. Elgendy and A. Elragal, "Big data analytics: a literature review paper," in Advances in Data Mining. Applications and Theoretical Aspects, vol. 8557 of Lecture Notes in Computer Science, pp. 214– 227, Springer, Cham, Switzerland, 2014.
- [29] H. C. Koh, W. C. Tan, and C. P. Goh, "A two-step method to construct credit scoring models with data mining techniques," International Journal of Business and Information, vol. 1, no. 1, pp. 96–118, 2006.
- [30] E. Kambal, I. Osman, M. Taha, N. Mohammed, and S. Mohammed, "Credit scoring using data mining techniques with particular reference to Sudanese banks," in Proceedings of the 1st IEEE International Conference on Computing, Electrical and Electronics Engineering (ICCEEE '13), pp. 378–383, August 2013.
- [31] Q. Liu, J. Wan, and K. Zhou, "Cloud manufacturing service system for industrial-cluster-oriented application," Journal of Internet Technology, vol. 15, no. 3, pp. 373–380, 2014.
- [32] D. Maaß, M. Spruit, and P. de Waal, "Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques," Decision Analytics, vol. 1, no. 1, pp. 1–17, 2014.
- [33] X. F. Du, S. C. H. Leung, J. L. Zhang, and K. K. Lai, "Demand forecasting of perishable farm products using support vector machine," International Journal of Systems Science, vol. 44, no. 3, pp. 556–567, 2013.
- [34] C.-J. Lu and Y.-W. Wang, "Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting," International Journal of Production Economics, vol. 128, no. 2, pp. 603–613, 2010.
- [35] H. Lee, S. G. Kim, H.-W. Park, and P. Kang, "Pre-launch new product demand forecasting using the Bass model: a statistical and machine learning-based approach," Technological Forecasting and Social Change, vol. 86, pp. 49–64, 2013
- [36] Chen, Deng, & Wan, 2015, Data Mining for the Internet of Things: Literature Review and Challenges , Hindawi Publishing Corporation International Journal of Distributed Sensor Networks
- [37] Dr. Shalini Bhaskar Bajaj Tushar Taneja, Aman Jatain. Predictive analytics on iot. In IEEE, pages 1312–1317. IEEE, 2009
- [38] Brijesh Iyer Prachi Deshpande. Research directions in the internet of every things(ioet). In International Conference on Computing, Communication and Automation (ICCCA2017), pages 1353–1357. IEEE, 2017



**Impact Factor:  
7.512**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details