A review on performance analysis of data mining methods in IoT

Ashutosh Kumar Dubey[*], Dimple Kapoor and Vijaita Kashyap
Chitkara University Institute of Engineering and Technology, Chitkara University
Punjab, India

Abstract
IoT is capable and helpful in interdisciplinary areas along with the convergence of multiple technologies and platforms. This study adheres the adaptation of data mining technologies along with big data and cloud computing with the IoT system with detailed discussion. This paper supports and provide systematic review and analysis based on the computational parameters and performance analysis. The analysis and discussion are based on the communication capability, system component communication, aspects of data mining, big data and cloud computing in IoT. Different types of transmission and communication barriers have also been discussed and analyze. Finally, based on the study and analysis a framework has been suggested for the smooth functioning of the IoT protocols.

Keywords
IoT, Data mining, Big data, Cloud computing and Computation capability.

1.Introduction
The interrelated network of different objects which behave smartly in the interconnection like sensors, mobile phones and radio-frequency identification (RFID) normally called Internet of Things (IoT) [1−3]. In terms of IoT internet plays an important role in communication and things belongs to the objects for the message transmission or simply the communication [3, 4]. It has been helpful and can be accommodated with Big Data, Cloud computing, biomedical devices, data delivery techniques, etc. [5−7]. The application areas are medical diagnosis, air pollution monitoring, temperature monitoring, security, network security etc. It is clear from these trends that the system generates an enormous amount of data. So, there is the need of pattern extraction, data analysis along with the techniques for matching and reasoning of data. It explores the need of data mining techniques. Data mining methods are capable in the extraction of patterns, relevant sets and establish an association and relationship [8, 9]. The need of integration among with the social and technical aspects between different objects is the current demand of IoT [10, 11].

There are several areas which can be integrated by the IoT are intelligent transportation, e-health, e-learning and industrial manufacturing with logistics [12]. It is discussed earlier that the main purpose is the interaction among different objects. The three major laying aspects are platform, architecture and operating systems. It is not necessarily same for all the objects [13]. In terms of data communication and data sharing IoT has application layer, network layer and perceptual layer. The main working principle and the responsibilities of these layers are different. Perception layer is mainly responsible for the object sensing, reading and identification. The examples are sensors, RFID tags and bar code readers. For the reading and collecting data from the perception layer network layer is responsible. The bridge between it and industries have been provided by the application layer.

The objectives to focus the current aspects in this area are as under:
1. To report and study the methods in the direction of system component communication capability in terms of IoT.
2. To discuss the system independence in terms of system component communication.
3. To discuss and analyses the IoT aspects in terms of data mining, cloud computing and Big Data.

---

*Author for correspondence

4. To discuss and analyses the IoT system in terms of reducing message loss and delay.
5. To discuss an efficient framework for the ease of applicability of different protocols in terms of methods adaptability, interoperability and security.

2.Related work

In 2018, Şimşek and Özdemir [14] discussed about the increasing data due to the IoT development. They have suggested that the real time analysis of the data is important. They have proposed a complex event processing method for this purpose. They have suggested the need of thorough knowledge of the data, its relations between data sources for the complex event processing rules. They have suggested that the rule imposing becomes typical if the quantity of data is increased. So, they have suggested the need of automatic extraction of rules. They have proposed a model for the rule's extraction. The rule extraction has been done from the unlabeled data. It has been done by clustering and rule mining algorithms. The model is evaluated in terms of classification performance and the results show that the proposed model is a promising solution for extracting complex event processing rules.

In 2018, Akulwar et al. [15] discussed the need of data mining. Data mining techniques have been suggested. It can be useful in the analysis of the educational purposes. According to the authors data mining techniques are capable in the analysis of the student's performance, mindsets, forecasting and in grouping. They have focused on different data mining techniques which are useful for the prediction of the student performance.

In 2018, Singh et al. [16] discussed about the digital transformation. They have developed self-advancing Goaf edge support systems (SAGES) based on IoT for the under-ground coal mines. It acts as the mobile roof support systems. They have explored the development of SAGES based on the IoT. They have suggested the probability of cyber-attacks in the mining industry. So, they have suggested the need of robust communication channel using datalogger. This is helpful in safety and reliability sending to the clouds. The information is reliably accessed.

In 2018, Jan and Sajeev [17] discussed about the learning model. They have proposed an intelligent model. It is based on the machine learning. They have named it an adaboosted semiparametric probabilistic neural network (PNN). It has been used for the classification of the IoT data. The

environment considered here is heterogeneous and dynamic. Prioritize resources can be obtained based on the classification. Critical services have been the main capabilities. Their results show that by their approach the computational cost is reduced.

In 2018, Alzahrani [18] discussed about the microblog's sentiment analysis. This analysis is based on the people's attitudes, their appraisals and their emotions. They have proposed an IoT mining machine. It has been used for the analysis based on Twitter. Twitter's API has been used for the harvesting purpose. They have developed a mining engine. It has been developed on the Raspberry Pi single-board microcomputer. They have programmed the IoT device. Sentiment analysis and opinion mining-based programming has been done. It has been used for the naive Bayes classifier. Positive or negative classification trends has been used. They have achieved 0.992 of accuracy from SemEval 2017. They have aggregated the sentiments of tweets in terms of the visualized graphs. Their experimental results show the feasibility and effectiveness of the IoT system.

In 2018, Lysogor et al. [19] discussed about the data interchange formats in terms of IoT for the heterogeneous networks. They have discussed about the data transfer issues. In terms of transmitting of the data the authors suggested the main source. According to them it is the satellite networks. Their research and analysis show that the iridium short burst data (SBD) are the better technological option for the IoT applications. But it has size limitations. Heterogeneous Iridium-LoRAWAN prototype has been developed along with the simulation model. It has been used for the comparison of different data exchange formats. Their results show that they have achieved four times higher data transfer.

In 2018, Shakir and Rakesh [20] discussed about air pollution and its adverse effect. So, they have suggested the need of air pollution, proper detection and analysis based on these detections. The pollutant proportionality finding is the ultimate goal. They have also estimated the effects of environmental parameters. The estimation is done through the WEKA tool.

In 2018, Monelli and Sriramoju [21] discussed about the data extraction through data mining. They have discussed different issues and the issues for developing the processing tools.

In 2018, Ibrahim and Shafiq [22] discussed about the massive data generated every day. The main source suggested by the authors is the social networks. It is in terms of IoT for the data generation. Time and space issues are the major concern in the above scenario as per the authors. They have presented a solution for the advanced and sophisticated distributed systems. They have used different training datasets. It has been used for the evaluation of the traditional pattern and their MapReduce solution. Their results in terms of performance support the aspects suggested by the authors.

In 2018, Subasi et al. [23] discussed the developments in information and communication. They have suggested wearable sensor technology as the main aspect in health monitoring in terms of IoT. They have presented data mining approach which is user dependency based. It has been used for the classification. They have achieved 99.89 % accuracy.

Their results are better in terms of robustness and reliability.

3.Method based analysis
Table 1 shows the analysis based on data mining, healthcare and IoT. It mainly covers the approaches based on data mining and health information corresponding to the IoT systems. The table clearly shows the method with the approach exploration along with the limitations. The main limitations found are the compositional feasibility along with the parametric evaluations. *Table 2* shows the analysis based on cloud computing, Big Data and IoT. It mainly covers the approaches based on cloud computing and Big Data corresponding to the IoT systems. The main limitations found are the practical implications along with the real time data streaming.

Table 1 Analysis based on Data Mining, Healthcare and IoT

S.No	Reference	Method	Approach	Limitations
1	[24]	Performance analysis in IoT through data mining techniques	They have discussed classification, clustering, and association-based mining as the data mining techniques focusing the applications also.	Parametric and computational feasibility have not been discussed.
2	[25]	Genetic algorithm	They have proposed a new way for the association rule mining based on genetic algorithm and multi-objective evaluation. It has been considered in terms of the bakery database.	They have suggested that in future support and confidence parameters may be added for the betterment.
3	[26]	IoT in case data streaming	The main focus considered by the authors is the incoming data stream. Problem approximation has been considered in case of data stream. It is based on the real time monitoring and analyzing in case of objects and the devices connected to the IoT.	Other aspects have not been included which may affect the performance.
4	[27]	Data mining and IoT	Frequent item set mining in case of IoT devices and for the data threshold have been discussed and analyzed.	Another dataset can be considered.
5	[28]	Fog computing for the prediction of forest fire	They have used support vector machine for the live sensor data. They have used thermal data, relative dampness, rain and wind velocity as the sensor input.	Other aspects have not been included which may affect the performance.
6	[29]	Opinion mining	The authors have discussed the implications of machine learning model in case of IoT system. Positive or negative sentiments have been considered based on the tweets. They have considered the authorized IPL hashtags for the experimentation.	Different ensemble techniques have been suggested by the authors for the future prospective.

Table 2 Analysis based on cloud computing, big data and IoT

S.No	Reference	Method	Approach	Limitations
1	[30]	Big data analysis tools in IoT	They have investigated the major data challenges in terms of IoT. They have suggested the use of Big Data to draw the data and explores the research challenges.	Practical implications are need to be explored.
2	[31]	Big data processing systems	The performance factor has been analyzed. The basis of this is the NoSQL performance which is mainly the MongoDB. They have used YCSB benchmark. Results based on NoSQL shows the capability of good performance advantage over SQL.	Other factors can be included to check the performance.

S.No	Reference	Method	Approach	Limitations
3	[32]	Veracity of IoT generated Big Data	They have proposed for the future development of IoT data. This has been used for the distributed nature of IoT data.	They have suggested to focus on the real time data streaming and latency minimization.
4	[33]	Decision oriented applications	Their approach discussion profound and explores the impact of the IoT. It is based on the big data sentiment analysis. The above platform has been applied for the big data sentiment analysis. It has been applied for the decision-oriented applications. The authors suggested the need of data analytics as the data generation is very high from the internet. They have discussed the big data sentiment analysis. Their suggestion supports analysis projects based on big data.	Practical implications are need to be explored.
5	[34]	Location-based smart shopping system	As per the authors localization has been found crucial in case of IoT. They have suggested the influenced impact and role of the data mining in case big data has been combined with the IoT. Smart shopping platform have been designed with the flexibility of location and other components. Then they have developed" location orbital" concept. It is capable in the mobile object's estimation. Then they have evaluated the performance. Their results show better precision in comparison to the other related trends.	Other factors can be included to check the performance.
6	[35]	IoT big data cluster analysis	They have suggested that the traditional k-means algorithm need some optimization enhancement for the suitability in case of IoT with Big Data. They have implemented k-means clustering with the Hadoop cloud clustering platform. They have selected center point for RFID based on the traditional clustering algorithm.	There is the need of efficient method for pattern identification and recognition.

Several new aspects were already been discussed in several literature which provides the hybridization of big data, data mining and IoT aspects. Some of the latest trends were discussed below showing the multivariant aspect in this direction.

In 2019, Hariri et al. [36] discussed about big data analytics. They suggested the uncertainty in the data collection from the sources like sensors, social media and financial records due to the noise, incompleteness and inconsistency. They mainly focused on the field of uncertainty and artificial intelligence techniques. They suggested the direction in terms of big data analytics aspect in terms of recognizing and mitigating uncertainty. In 2019, Song et al. [37] discussed about the IoT based medical services. They proposed medical big-data mining process development. Their proposed approach follows the guidelines of the knowledge discovery in big-data process for data mining and topic modelling. Big data was used by the open system of the health- and medical-services. They also conducted topic-modeling accuracy evaluation. In 2019, Atmani et al. [38] discussed about the connection of IoT in terms of wearable devices.

They suggested big data technologies for the user experience automation and the massive amount of data generation. They reviewed different aspects in the direction of IoT in the big data aspects. They provide a collaborative discussion in terms of IoT frameworks and platforms. In 2019, Hajjaji et al. [39] discussed about the big data and IoT based applications in smart environments. They considered papers from 2011 to 2019. It consists of identification, screening, eligibility, and inclusion. Their approach indicates that the combination of big data and IoT may be helpful in different smart environment applications. In 2020, Moghadas et al. [40] discussed about IoT. They proposed a system for health monitoring system. They used AD8232 sensor module and Arduino's electronic board. For classification they used k-nearest neighbor. It was used for the classification and validation of cardiac arrhythmia. In 2020, Sankaranarayanan et al. [41] discussed the aspects of IoT in different areas like smart city, oil mining and transportation. Their main focus is the integration of distributed deep learning and data flow. They proposed a novel data flow and distributed deep neural network based IoT edge model for big data environment. They achieved 33% result in latency reduction.

4.Problem statements

The following problems as the intervening have been found corresponding to the study investigated above with the literature exploration.

1. There is the need of interoperability in the objects to improve the IoT system performance.
2. There is the need of reducing technical specification in terms of data communication.
3. For the resource constraint devices like IoT there is the need of selection of appropriate lightweight protocols.
4. There is the need of working in terms of performance parameters like delay, communication platform, message loss and other system performance enhancer.
5. There is the need of security adaptation along with the detection of attacks if any, in the way the communication is being held.

5.Suggested framework

Figure 1 shows the suggested hybrid framework which is based on the role of data mining, big data, cloud computing along with the IoT interface. This framework supports the communication gateway with the advantages of other computing platforms with the support of communication resources. It also supports the cloud service model like IaaS, PaaS and SaaS with message transformation capability. Data mining handles the pattern extraction along with the visualization and other aspects for better data compatibility to provide IoT system handling with the lightweight protocols. *Figure 2* shows the applicability of suggested hybrid framework in different domain like computing, automation, health, visualization, transportation, network and support in different devices.
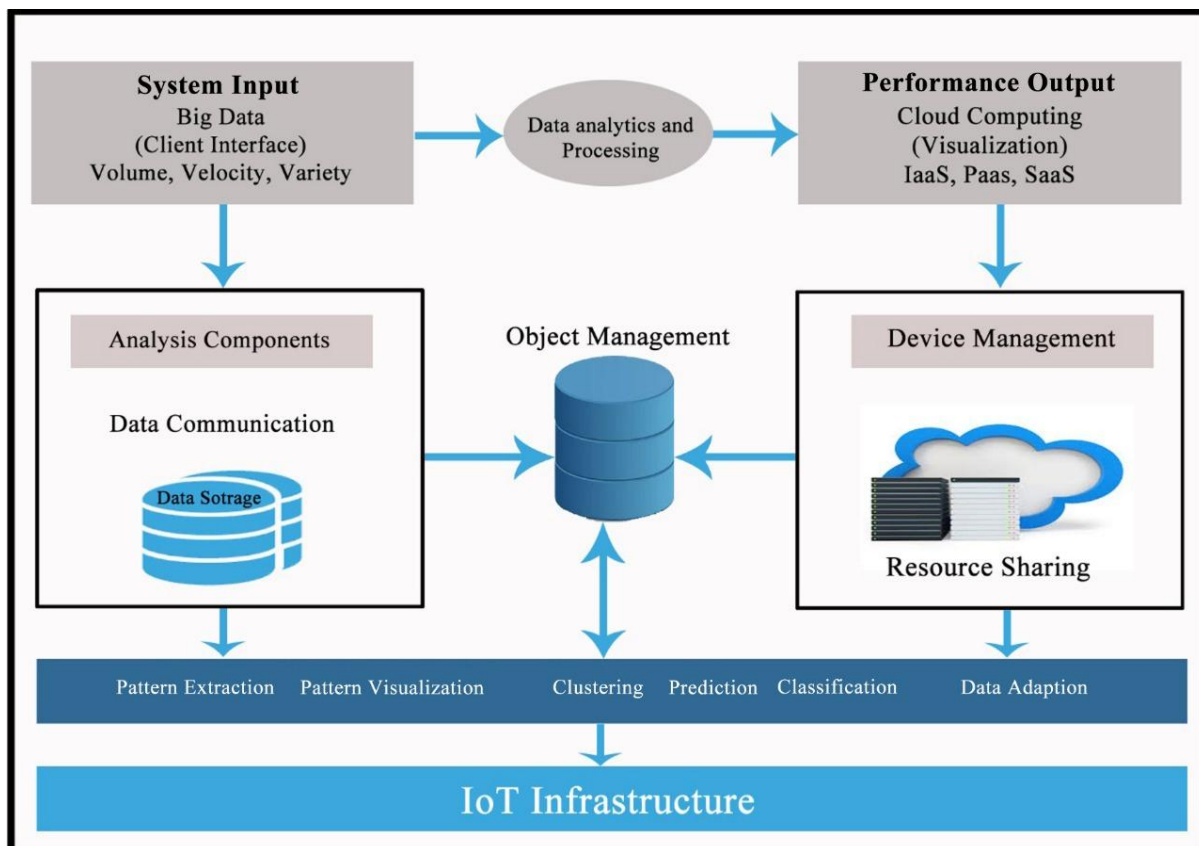

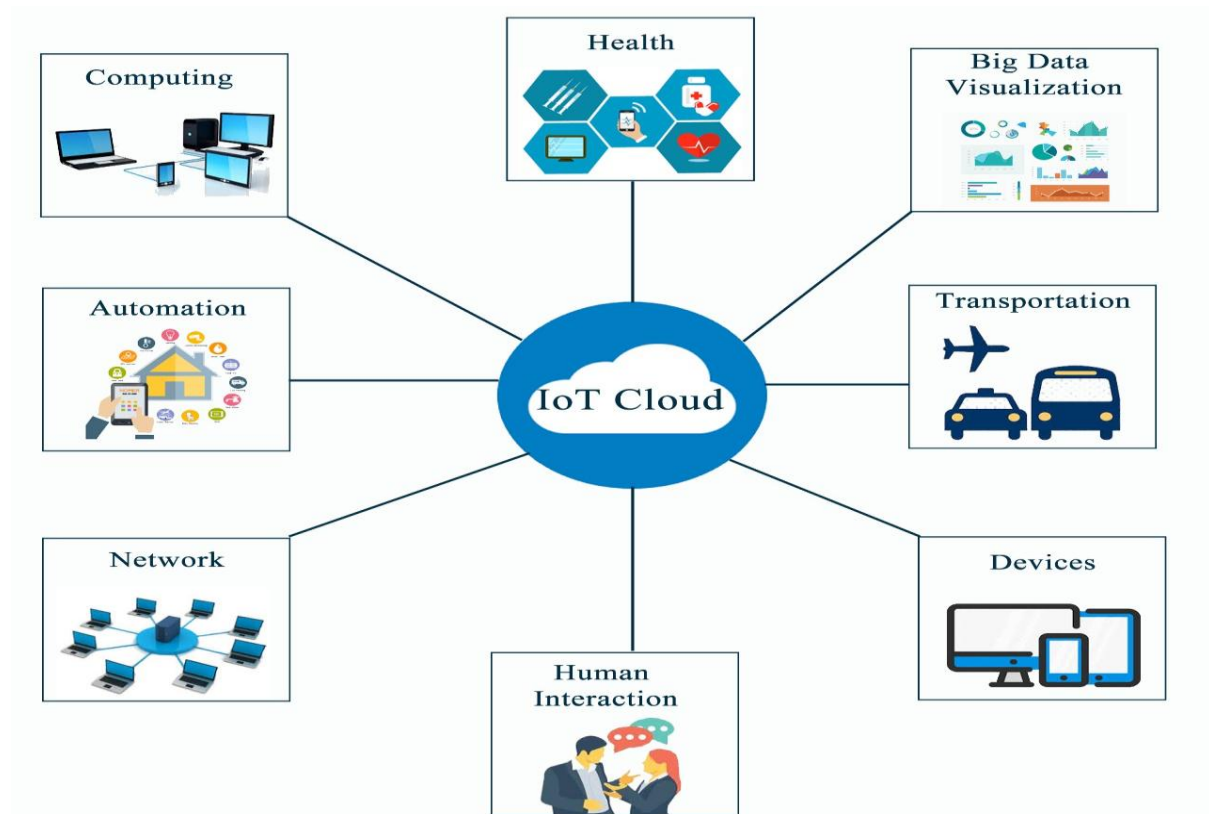**Figure 1** Suggested hybrid framework for data computation and analysis

Figure 2 Applicability of suggested hybrid framework in different domain

6.Conclusion and future work

This paper explores the possibility of the hybridization of data mining, Big data and cloud computing along with the interfacing of IoT based system. For this purpose, analysis based on the study has been presented and discussed. First part covers the overall literature. In the second part the analysis presented based on data mining and IoT and then the combination of cloud computing with Big data have been presented. Based on the discussion and analysis advantages and gaps have been explored. Finally, future framework has been suggested for the implementation and validation purpose.

Based on the review and analysis the future suggestions are as follows:

1. There is the need of efficient method for pattern identification and recognition along with the influence analysis of the factors associated.
2. There is the need of emphasis on the real time data analysis.
3. There is the need of performance analysis in terms of communication and computation.

References
[1] Stankovic JA. Research directions for the internet of things. IEEE Internet of Things Journal. 2014; 1(1):3-9.
[2] Andreev S, Galinina O, Pyattaev A, Gerasimenko M, Tirronen T, Torsner J, et al. Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap. IEEE Communications Magazine. 2015; 53(9):32-40.
[3] Miorandi D, Sicari S, De Pellegrini F, Chlamtac I. Internet of things: vision, applications and research challenges. Ad Hoc Networks. 2012; 10(7):1497-516.
[4] Sriram RD, Sheth A. Internet of things perspectives. IT Professional. 2015; 17(3):60-3.
[5] Patel P, Patel HB, Shrimali B. Data delivery techniques in content centric routing for IoT: a systematic review. International Journal of Advanced Technology and Engineering Exploration. 2018; 5(48):445.
[6] Mishra A, Mohapatro M. An IoT framework for bio-medical sensor data acquisition and machine learning

for early detection. International Journal of Advanced Technology and Engineering Exploration. 2019; 6(54):112-25.

[7] Dubey AK, Dubey AK, Agarwal V, Khandagre Y. Knowledge discovery with a subset-superset approach for mining heterogeneous data with dynamic support. In CSI sixth international conference on software engineering 2012 (pp. 1-6). IEEE.

[8] Mohamed MH, Waguih HM. A proposed academic advisor model based on data mining classification techniques. International Journal of Advanced Computer Research. 2018; 8(36):129-36.

[9] Dubey AK, Gupta U, Jain S. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. International Journal of Computer Assisted Radiology and Surgery. 2016; 11(11):2033-47.

[10] Atzori L, Iera A, Morabito G. The internet of things: a survey. Computer Networks. 2010; 54(15):2787-805.

[11] Garcia-Morchon O, Falck T, Heer T, Wehrle K. Security for pervasive medical sensor networks. In annual international mobile and ubiquitous systems: networking & services, mobiquitous 2009 (pp. 1-10). IEEE.

[12] Da Xu L, He W, Li S. Internet of things in industries: A survey. IEEE Transactions on Industrial Informatics. 2014; 10(4):2233-43.

[13] Oen HM. Interoperability at the Application Layer in the Internet of Things (Master's thesis, NTNU).

[14] Şimşek MU, Özdemir S. CEP rule extraction from unlabeled data in IoT. In international conference on computer science and engineering 2018 (pp. 429-33). IEEE.

[15] Akulwar P, Pardeshi S, Kamble A. Survey on different data mining techniques for prediction. In international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) 2018 (pp. 513-9). IEEE.

[16] Singh A, Singh UK, Kumar D. IoT in mining for sensing, monitoring and prediction of underground mines roof support. In international conference on recent advances in information technology 2018 (pp. 1-5). IEEE.

[17] Jan T, Sajeev AS. Boosted probabilistic neural network for IoT data classification. In international conference on dependable, autonomic and secure computing, international conference on pervasive intelligence and computing, international conference on big data intelligence and computing and cyber science and technology congress 2018 (pp. 408-11). IEEE.

[18] Alzahrani SM. Development of IoT mining machine for Twitter sentiment analysis: mining in the cloud and results on the mirror. In learning and technology conference 2018 (pp. 86-95). IEEE.

[19] Lysogor II, Voskov LS, Efremov SG. Survey of data exchange formats for heterogeneous LPWAN-satellite IoT networks. In Moscow workshop on electronic and networking technologies 2018 (pp. 1-5). IEEE.

[20] Shakir M, Rakesh N. Investigation on air pollutant data sets using data mining tool. In international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) 2018 (pp. 480-5). IEEE.

[21] Monelli A, Sriramoju SB. An overview of the challenges and applications towards Web mining. In international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) 2018 (pp. 127-31). IEEE.

[22] Ibrahim R, Shafiq MO. Towards a new approach to empower periodic pattern mining for massive data using map-reduce. In international conference on big data 2018 (pp. 2206-15). IEEE.

[23] Subasi A, Radhwan M, Kurdi R, Khateeb K. IoT based mobile healthcare system for human activity recognition. In learning and technology conference 2018 (pp. 29-34). IEEE.

[24] Batra I, Verma S, Janjua K. Performance analysis of data mining techniques in IoT. In international conference on computing sciences 2018 (pp. 194-9). IEEE.

[25] Bari T, Parvat T, Badodekar R. Novel genetic algorithm for association rule mining with multi-objective extraction for bakery database. In international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) 2018 (pp. 755-9). IEEE.

[26] Alieksieiev V. One approach of approximation for incoming data stream in IoT based monitoring system. In second international conference on data stream mining & processing 2018 (pp. 94-7). IEEE.

[27] Wang Z, Liang W, Zhang Y, Wang J, Tao J, Chen C, Yan H, Men T. Data mining in IoT era: a method based on improved frequent items mining algorithm. In international conference on big data and information analytics 2019 (pp. 120-5). IEEE.

[28] Aakash RS, Nishanth M, Rajageethan R, Rao R, Ezhilarasie R. Data mining approach to predict forest fire using fog computing. In second international conference on intelligent computing and control systems 2018 (pp. 1582-7). IEEE.

[29] Dubey KP, Agrawal S. An opinion mining for Indian premier league using machine learning techniques. In international conference on internet of things: smart innovation and usages (IoT-SIU) 2019 (pp. 1-4). IEEE.

[30] Nicolalde FC, Silva F, Herrera B, Pereira A. Big data analysis tools in IoT and their challenges in open researches. In Iberian conference on information systems and technologies 2018 (pp. 1-6). IEEE.

[31] Kaur MJ, Maheshwari P. Comparison study of big data processing systems for IoT cloud environment. In HCT information technology trends 2018 (pp. 104-8). IEEE.

[32] Liu X, Tamminen S, Su X, Siirtola P, Röning J, Riekki J, Kiljander J, Soininen JP. Enhancing veracity

of IoT generated big data in decision making. In international conference on pervasive computing and communications workshops 2018 (pp. 149-54). IEEE.

[33] Dwivedi A, Pant RP, Pandey S, Kumar K. Internet of things'(IoT's) impact on decision oriented applications of big data sentiment analysis. In international conference on internet of things: smart innovation and usages 2018 (pp. 1-10). IEEE.

[34] Rezazadeh J, Sandrasegaran K, Kong X. A location-based smart shopping system with IoT technology. In world forum on internet of things 2018 (pp. 748-53). IEEE.

[35] Bin N. Research on methods and techniques for IoT big data cluster analysis. In international conference on information systems and computer aided education 2018 (pp. 184-8). IEEE.

[36] Hariri RH, Fredericks EM, Bowers KM. Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data. 2019; 6(1):44.

[37] Song CW, Jung H, Chung K. Development of a medical big-data mining process using topic modeling. Cluster Computing. 2019; 22(1):1949-58.

[38] Atmani A, Kandrouch I, Hmina N, Chaoui H. Big data for internet of things: a survey on IoT frameworks and platforms. In international conference on artificial intelligence and symbolic computation 2019 (pp. 59-67). Springer, Cham.

[39] Hajjaji Y, Boulila W, Farah IR, Romdhani I, Hussain A. Big data and IoT-based applications in smart environments: a systematic review. Computer Science Review.;39:100318.

[40] Moghadas E, Rezazadeh J, Farahbakhsh R. An IoT patient monitoring based on fog computing and data mining: cardiac arrhythmia usecase. Internet of Things. 2020; 11:100251.

[41] Sankaranarayanan S, Rodrigues JJ, Sugumaran V, Kozlov S. Data flow and distributed deep neural network based low latency IoT-edge computation model for big data environment. Engineering Applications of Artificial Intelligence. 2020; 94:103785.

Dr. Ashutosh Kumar Dubey received his PhD degree in Computer Science and Engineering from JK Lakshmipat University, Jaipur, Rajasthan, India. He is currently in the department of Computer Science and Engineering, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India. He is Senior Member of IEEE and ACM. He has more than 13 years of teaching experience. He has authored a book name Database Management Concepts. He has been associated with many international and national conferences as the Technical Program Committee member. He is also associated as the Editorial Board Member/ Reviewer of many peer-reviewed journals including Elsevier, Springer, BMJ, IOS Press, Bentham Science, Thieme Publishing Group, etc. His research areas are Data Mining, Optimization, Machine Learning, Cloud Computing, Artificial Intelligence and Object-Oriented Programming.
Email: ashutosh.dubey@chitkara.edu.in

Dimple Kapoor received her Master in Technology degrees in Computer Science and Engineering from Meerut Institute of Engineering & Technology, Meerut, U.P in 2015. She is currently an Assistant Professor in the Department of Computer Science and Engineering at Chitkara University, Rajpura, Punjab. Her research interests include Image Processing, Data Mining and IOT.
Email: dimple.kapoor@chitkara.edu.in
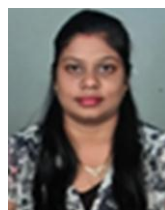
Vijaita Kashyap received her Master in Technology degrees in Computer Science and Engineering from Samrat Ashok Technological University, Vidisha, M.P in 2012. She is currently an Assistant Professor in the Department of Computer Science and Engineering at Chitkara University, Rajpura, Punjab. Her research interests include Data Mining and IOT.
Email: vijaita.kashyap@chitkara.edu.in