Journal of Engineering Research

Volume 6 | Issue 5 Article 29

2022

Data Mining in Internet of Things Systems: A Literature Review

Amany M. Sarhan

Vol. 6, No. 5 – 2022 Journal of Engineering Research (ERJ)

Data Mining in Internet of Things Systems: A Literature Review

Amany M. Sarhan

Computer and Control Engineering Department, Faculty of Engineering, Tanta University, Egypt
Emails: amany_sarhanr@f-eng.tanta.edu.eg

***Abstract*- The Internet of Things (IoT) and cloud technologies have been the main focus of recent research, allowing for the accumulation of a vast amount of data generated from this diverse environment. These data include without any doubt priceless knowledge if could correctly discovered and correlated in an efficient manner. Data mining algorithms can be applied to the Internet of Things (IoT) to extract hidden information from the massive amounts of data that are generated by IoT and are thought to have high business value. In this paper, the most important data mining approaches covering classification, clustering, association analysis, time series analysis, and outlier analysis from the knowledge will be covered. Additionally, a survey of recent work in in this direction is included.**

Keywords: **Internet of Things; Cloud computing; Data mining, Classification, Clustering, Outlier detection, K-nearest Neighbour, K-means, Support Vector Machine.**

I. INTRODUCTION

The fundamental idea of the Internet of Things is that objects and things can be linked to the internet to enable them to communicate with humans and with each other, where each device has a distinct identity, and be automatically recognized [53]. Therefore, it can be stated that in the Internet of Things (IoT), the internet can be viewed as a global platform that enables machines and intelligent objects to communicate, compute, make decisions, and coordinate with humans globally [7, 9]. We can see that the number of IoT devices are growing even in a faster way than human. By the year 2020, "50 billion small embedded sensors and actuators will be connected to the Internet, and the Internet of Things will create 14.4 trillion dollars of value at stake for industries in the next ten years," according to Ericsson and Cisco [23]. We can therefore state that the Internet of Things environment will be made up of a very large number of connected heterogeneous devices.

Agriculture [6, 9, 63], public infrastructure [4], catastrophe management, education and fear, electricity, environment, health, transport, and mobility [41, 50, 54] are just a few of the data collection fields. An IoT decision making application should be able to retrieve historical and real-time data from a variety of sources in a number of ways, including big data, conventional data sources, and user personal information [50, 54].

Large volumes of data should be analyzed as the Internet of Things (IoT) expands, and the most recent algorithms should be modified to work with big data. Additionally, this generated knowledge will be essential for making better decisions, enhancing system performance, and providing the best possible management of resources and services. The selection or synthesis of the best data mining algorithm is a difficult task in any IoT-enabled smart environment. Such an algorithm should generate insightful analytics, accurately forecast future events, and manage the network and services effectively while adhering to all restrictions. Unique identification is essential from the perspective of IoT data mining as well. Better actuation control can be achieved through strong knowledge derived from a better understanding of infrastructure-related data in order to cope with real-time challenges. Another significant challenges in the field are collecting, storing, and managing the large number of devices along with their associated features. In this paper, a deep look on the data mining for the IoT platforms will be given concentrating on real applications found in the literature [20, 54].

II. INTERNET OF THINGS MODELS

Sensor technology, radio frequency identification technology, and embedded computer technology are the three crucial innovations used in the real-world implementation of the Internet of Things [6, 45]. The term "sensor" refers to parts or tools that can gather data from a sensing object and transform it into a useful information type by applying rules or methods. For the purpose of defining specific objectives through radio signals and reading and writing specific target data, identification of the radio frequency combines radio frequency technologies with embedded technology [3]. The technology is made up of integrated circuit technology, sensor technology, computer software, and electronic application technology. The embedded system, which is made up of a CPU, storage, I/O, and applications and is portable and highly stable, manages computing power and carries out all of the functions of the application device [40].

Numerous studies have been conducted in this area, including those on the flagship, mobile devices, web-related data, and social data, as a result of the growing demand for intelligent and integrated communities [28]. The many fields in which it has been applied—including agriculture, civil infrastructure, disaster relief and response, planning, education, electricity, the effectiveness of the atmosphere, health and wellbeing, including health and human services—have produced an abundance of knowledge [6, 41, 44]. For those who have less physical contact, the use of technical tools to identify people is the preferred method. Internet and bank ATM passwords are crucial [42]. Despite the importance of comprehending, interpreting, and processing data, politicians, municipal administrators, citizens, and other individuals can greatly profit from the ability to view data in real-time.

IoT's primary building blocks include: hardware, which consists of sensors; middleware, which facilitates communication between components; data handling; and processing and visualization of data [29, 38]. According to a Cisco report [13], more than 14.7 billion Internet connections will be made for Internet of Things applications up until 2023. Everything and everyone can be a part of the Internet,

according to the IoT paradigm. The way people interact with one another and the objects around them is redefined by this vision [27]. By taking advantage of the opportunities provided by Internet technology, the recent adoption of various wireless technologies positions IoT as the following revolutionary technology [54]. IoT will develop into a brand-new class of infrastructure resource, similar to water, electricity, gas, and roads [6, 20].

The integration of data collection mechanisms into the system is the first step toward achieving an IoT-based system. Sensors are used to monitor and alert changes in environments such as temperature, weight, pressure, light conditions, noise levels, motion, humidity, and so on. Sensors are frequently embedded in objects like machines or devices. Sensors are tailored to precisely meet the specific application needs. All of these sensors provide data that can be used to detect dynamic patterns. Sensors may be wired or wireless. A sensor network connects sensors and sends signals between them. Bluetooth, ZigBee, Wi-Fi, RFID, DSL, LAN, LoRaWAN, LTE, 5G, and other technologies can establish communication between IoT devices. Wireless technologies reduce installation and maintenance costs, which is critical to the advancement of IoT. Sensor nodes transmit their findings to a small group of special nodes known as data sinks [66].

Figure 1 depicts the overall architecture of an IoT cloud-based system, which has a hierarchical structure of several layers. It includes a fog layer for fast response processes and a cloud layer for the storage of the overall data [7].

1- IoT sensors and actuators layer: The distributed sensor grid is built into infrastructure to collect and transmit real-time environmental and social data. Data acquisition elements are in charge of collecting and storing external data locally. It can record any type of data, including images, video, sound, temperature, humidity, pressure, and so on. Between the distributed sensor layer and the service-oriented middleware layer, network elements are used for data transfer and information routing. This layer is found in other studies [32] as two separate layers, sensors and network layer.
2. The fog devices layer which is used as intermediate layer between the IoT layer and the cloud layer. Most of the recent systems relies for this layer to enhance the latency of data movement to and from the IoT devices.
3. The cloud layer that contains several inlayers considered as service-oriented middleware layers which are in charge of large amounts of data storage, as well as real-time analysis and processing. It is built on cloud computing, distributed memory, and highly efficient index services. The findings can be used to aid decision making and the efficient operation of smart city applications. The data mining and knowledge discovery processes are part of the cloud layer tasks that can be represented as direct reports for the users or used for enhancing the overall user/system experience. However, part of the data stored at the fog layer can be used for partial data mining and knowledge discovery processes on the local data found at this layer [30].
4. The IoT application layer for end users provides tailored intelligence services to various domains and interacts directly with the user. It presents information to the user in an understandable format, such as graphical forms, tables, or other presentation types, and facilitates interaction with the system.

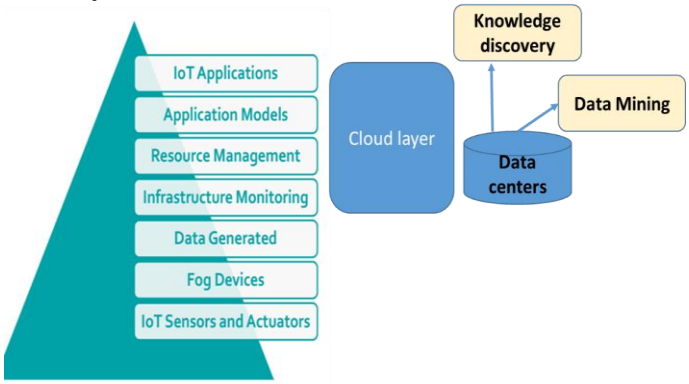

Fig. 1. Overall architecture of an IoT cloud-based system

III. DATA MINING TECHNIQUES

Data mining is the process of extracting hidden information using algorithms while also identifying new, intriguing, and potentially practical patterns in sizable datasets. Information is gathered through knowledge extraction, data/pattern analysis, and database discovery (or "mining") [42]. Any data mining process's main goal is to create a predictive or descriptive model of a large amount of data that is both effective and able to generalize to new data [49]. Data mining is the process of locating relevant information from vast the volume of data kept in databases, data warehouses, or other information storage facilities.

DM is a broad field that includes many algorithms and techniques for extracting information from data, ranging from statistics to machine learning and information theory [5]. The goal of data mining is to create computer programs that extract hidden, previously unknown, and potentially valuable information from data. The process must be automated or, more commonly, semi-automated, and the regularities or patterns discovered must be useful. Under the characteristics of big data, big DM must extend the entire process to the front and back end. This entails processing and analyzing large amounts of heterogeneous data, automatically discovering and extracting implicit, hidden patterns, rules, and knowledge, and visualizing them in an understandable manner [16].

Today, data mining can be automated reinforcement learning, unsupervised learning, or both. When conducted in a hierarchy of layers, computer-assisted learning becomes more precise. Machine learning is the term for this automatic feature extraction through supervised or unsupervised learning in a hierarchical fashion (ML) [6, 41, 44].

Not only is data mining essential for knowledge discovery, but it is also crucial for removing internet unimportant data that can harm the meaningful data. Ganz et al. [21] suggested "data abstraction" as a suitable technique. They examined various abstraction techniques and made the suggestion that data centers should only store abstracted results rather than the original data. An essential component of knowledge discovery is data mining.

Generally speaking, data mining process consists of the following steps as shown in fig. 2:

1- ***Data preparation***: in which data is prepared for mining. It includes various processes like data integration from various data sources (as in IoT devices), removal of

noise from data, and preprocessing the data to make it easier to mine (like feature selection and extraction, transform raw data into the proper format for analysis, normalization, dimension reduction, etc).

2- ***Data mining***: Applying algorithms on data in order to find the possible patterns and knowledge at a higher level.

3- ***Data visualization***: show the user the knowledge that has been mined from the data.

DL gave the advantage of merging both data pre-processing and data mining units in one process. Additionally, the output of DL is assessed and represented in knowledge that can be used by both machines and people, which is then used by the IoT infrastructure.

A. Classification

Classification is one of the most well-known and widely used DM technique [43]. Classification algorithms seek functions and models to determine which of a set of categories (sub-populations) a new observation belongs to, based on a training set of data containing observations (or instances) whose category membership is known. Classification is a supervised learning paradigm. The training or construction phase, in which the model learns from a given set of labeled samples, and the classification phase, in which the model outputs the label with the highest likelihood for a given sample, are the two main phases of supervised learning. It has been demonstrated that classifier performance typically improves as the volume of training data increases [39, 46].

Classification is defined as assigning objects to previously established classes. Its goal is to correctly predict the correct class for each data object. It is a supervised learning process because the targeted labels are presupposed to be known before processing. Before being used to classify unlabeled or unknown objects or data, the classifier must be trained, i.e, build the set of rules to work with in the classification phase. This can be done using labeled or known data items. The classifier is trained with the data of the same nature to identify, therefore, it is an application-based task.

The data is split into two categories: the training set and the testing set. Sometimes it could be split into three categories: the training set, the validation set and the testing set. The split should be random and covers the possible classes of the data. The classifier is first built using the training set, and then validated through the testing set. The classifier performance is judged using many metrics like accuracy, recall and precision.

A variety of classification models are available in the literature like: Decision Tree Induction, Bayesian classification, Rule-based classification, classification by backpropagation, Support Vector Machine, k-Nearest Neighbor, Deep Neural Network, and Ensemble methods. For difficult, extensive IoT Application problems, a set of classifiers can also implement the fusion of different classification techniques [20, 54].

A.1 Decision tree

Decision Trees (DT) are a popular technique that has proven to be effective for regression and classification. In the leaves, the DT predictive model represents observations in the branches and conclusions about the target value. The DT breaks down a data set into subsets by DTs in the first stage (construction), and then an associated DT is incrementally developed. The second stage involves pruning within the nodes and, if necessary, replacing the node with a regression plane. DT is one of the most widely used learning methods. It can handle mixed-type data and missing values, which are common in most data sets in a smart city application. Classification and Regression Trees are two methods for DT construction (CART), ID3, C4.5 (J48), and T3 [10, 24].

It is a flowchart-like tree structure in which internal nodes are represented by rectangles and leaf nodes by ovals. Every internal node have two or more child nodes. All internal nodes have splits that test the value of an attribute expression. Arcs connecting internal nodes to their children are labeled with test results. A class label is assigned to each leaf node. ID3, or Iterative Dichotomiser 3, is a straightforward decision tree learning algorithm. The C4.5 algorithm is an improved version of the ID3 algorithm that uses gain ratio as a splitting criterion [18]. The distinction between ID3 and C4.5 algorithms is that ID3 employs binary splits, whereas C4.5 employs multiway splits. SLIQ (Supervised Learning In Quest) can handle large data sets with ease and in less time. SPRINT (Scalable Parallelizable Induction of Decision Tree algorithm) is also fast and scalable, with no storage constraints on larger data sets [67]. Classification and Regression Trees (CART) is a decision tree algorithm that is nonparametric. Depending on whether the response variable is categorical or continuous, it generates classification or regression trees.

A.2 Random Forest

Random Forest (RF) is an ensemble learning model for classification and regression that works by training a batch of DT [51]. Cross-validation is used to select the output class, which is the mean prediction (in regression problems) or mode of the classes (in classification problems) of the ensemble. RF operates very efficiently on both scaled and unscaled data. It can handle data sets with unbalanced classes and generates strong predictive models while avoiding overfitting. RF is an appealing prediction technique due to two characteristics: its ability to achieve high prediction accuracy and its usability of desired capabilities, such as daily electricity data consumed by various appliances. These two characteristics distinguish RF as a distinct and desirable model for data analysis [36].

Fig. 2. Data mining process

A.3 Nearest Neighbor algorithm

The Nearest Neighbor algorithm, which is designed to find the nearest point of the observed object, introduced the KNN (K-Nearest Neighbor) algorithm. The KNN algorithm's main idea is to find the K-nearest points [17]. There are a lot of different improvements for the traditional KNN algorithm, such as the Wavelet Based K-Nearest Neighbor Partial Distance Search (WKPDS) algorithm, Equal-Average Nearest Neighbor Search (ENNS) algorithm, Equal-Average Equal-Norm Nearest Neighbors code word Search (EENNS) algorithm, the Equal-Average Equal-Variance Equal-Norm Nearest Neighbor Search (EEENNS) algorithm [37].

A.4 Bayesian networks

Bayesian networks are directed acyclic graphs with vertices represent Bayesian random variables. Conditional dependencies are represented by edges. Non-connected nodes represent variables that are conditionally independent of one another. A BN in data modeling creates implicit assumptions about variable dependencies, despite the fact that in the real world, two variables are never truly and fully independent. A Naive Bayes (NB) classifier is a subset of BN that employs Bayes' theorem in a naive manner by assuming that every predictor variable is conditionally independent on the class (i.e., attribute) label [22]. As a result, an NB is a straightforward stochastic classifier based on Bayes' theorem and strong independence assumptions. Brisimi et al. used the Likelihood Ratio Test in their work [64]. These classifiers, which are based on Bayesian networks, have many advantages, such as model interpretability and adaptability to complex data and classification problem settings. Naive Bayes, selective naive Bayes, seminaive Bayes, one-dependence Bayesian classifiers, K-dependence Bayesian classifiers, Bayesian network-augmented naive Bayes, unrestricted Bayesian classifiers, and Bayesian multinets are all investigated in [52].

A.5 Support Vector Machine

Support Vector Machine is one of the famous supervised learning models that analyzes data and recognizes patterns based on statistical learning. By mapping the input vectors into the high-dimensional feature space in an incredibly nonlinear manner, SVM creates a binary classifier, the so-called optimal separating hyperplanes. SVM is frequently used in many areas including pattern recognition, medical diagnosis, and text classification. Many versions of the algorithm are available, namely; GSVM (granular support vector machines), FSVM, TWSVMs, and VaR-SVM (value-at-risk support vector machines), and RSVM (ranking support vector machines) [46].

A. 6 Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is a well-known supervised learning technique. An ANN is a powerful nonlinear modeling tool that mimics the functioning of biological neurons. Training an ANN entails fine-tuning the network's weights and biases. The goal is to maximize network prediction performance by minimizing the difference between all network outputs and desired outputs or targets on validation data. The NN can learn relevant statistical information from a suitable amount of training data using the Back Propagation (BP) algorithm, and the mathematical information learned can reflect the function mapping relation of the input-output data model [41, 44]. Multilayer Perceptron, a Neural Network with a fully connected layer, is an earlier ANN architecture.

B. Clustering

A collection of related objects is referred to as a cluster. The clustering algorithm divides the collected objects into a predetermined number of clusters, each of which has objects with a similar set of characteristics. Data are divided into meaningful groups so that patterns within the same group are somewhat similar to patterns within other groups where group members differ in the same way. Clustering is an unsupervised learning technique, in contrast to classification, which depends on prior knowledge to direct the partitioning process [65]. Clustering will be helpful in this situation by grouping the objects into various groups for appropriate similar actions based on the features found in those objects. The division of points into various groups is a simple example of a clustering problem.

The clustering algorithms are based on finding centroids for the clusters and measure the least distance of an object to these centroids to assign it to this cluster. Usually, clustering techniques are designed for a specific application for which they perform well. Examples of clustering algorithms are: K-mean, K-Nearest Neighbor, K-medoids, hierarchical clustering (CURE, SVD, ROCK, BIRCH) [1], Density based clustering (DBSCAN, OPTICS, DENCLUE), Grid based clustering (STRING, WaveCluster [2, 56-60, 65]. With the development of sensor technologies, IoT and WSN create user environments that are intelligent enough to recognize user activity and respond appropriately.

(i) A hierarchy tree is created by subgroups splitting into larger, higher level groups, those higher level groups joining together, and so on. There are two types of hierarchical clustering techniques: agglomerative (bottom-up) and divisive (top-down). Beginning with one-point clusters, the agglomerative clustering recursively merges two or more clusters. Contrarily, clustering is a top-down method that begins with a single cluster that contains all of the data points and recursively divides it into suitable subclusters. Typical studies include CURE (Clustering Using Representatives) [57] and SVD (Singular Value Decomposition) [47].

(ii) Partitioning algorithms identify clusters either by identifying regions with a high concentration of data or by iteratively relocating points between subsets. Research on SNOB, MCLUST, k-medoids, and k-means is included in the related research [54]. The goal of density-based partitioning techniques is to find low-dimensional, densely connected data, also referred to as spatial data. DBSCAN (Density Based Spatial Clustering of Applications with Noise) [2] is one of the related studies. Hierarchical agglomeration is one of the processing phases used by grid-based partitioning algorithms, which also perform space segmentation.

(iii) Researchers precluster items or categorical attribute values in order to handle categorical data; typical research includes ROCK [58].

(iv) Scalable clustering research, such as DIGNET [56] and BIRCH [65], faces scalability issues for computing time and memory requirements.

(v) High dimensionality data clustering techniques like DFT [18] and MAFIA [26] are made to handle data with a large number of attributes.

C. Associations Analysis

Frequent patterns are data objects, sets of data objects, or sequences of events that recur frequently in a system [15, 33]. Exploiting these recurring patterns provides an excellent analytical understanding of the user's activity in a pleasant setting. Market basket analysis or transaction data analysis are the main foci of association rule mining, which aims to find rules that show attribute-value associations that are common and aid in the creation of more comprehensive and qualitative knowledge, both of which aid in decision-making [51]. By mining association, correlation, and other relationships among the data, pattern recognition plays a crucial role in promoting business maneuvers [10, 12] which has broad applications in market basket situations. Data will be processed sequentially for the first catalog of association analysis algorithms.

Order of events detected is also important in pattern mining. Sequential patterns are patterns that are observed in a specific order, and sequential pattern mining is the process of extracting those patterns. The first sequential mining problem was introduced by Aggrawal [10], and it was based on the sequence of transactions for customer purchases. As the sequence of events are frequently observed over a specific time period for activity discovery and recognition, MavHome, and GreaterTech smart home, sequential pattern mining is preferred over frequent pattern mining in the sensor environment [54].

There are numerous extension algorithms that have been used to find associations and then intratransaction associations using a priori based algorithms. It clusters into 2 types based on the format of the data record: Vertical and Horizontal Database Format Algorithms, Database format algorithms. Given large amounts of data, the pattern growth algorithm is more complicated but potentially faster to calculate. FP-Growth algorithm is a common algorithm [20].

In some cases, the data would be an event flow, so the challenge would be to identify event patterns that frequently occur together. The typical algorithm is PROWL [20]. It is divided into two categories: event-based algorithms and event-oriented algorithms. Various algorithms are created in order to benefit from distributed parallel computer systems, such as Par-CSP [20].

D. Time series Analysis

A time series is a grouping of temporal data items; time series data have large data sizes, high dimensionality, and continuous updating. Time series tasks frequently rely on three components: indexing, similarity measures, and representation. Reducing the dimension is one of the main goals of time series representation, which can be categorized into three types: model-based representation, non-data-adaptive representation, and data-adaptive representation. The parameters of the underlying model for a representation are sought after by model-based representations. The parameters of a transformation will alter in data adaptive representations in accordance with the data available and related works. Time series analysis typically uses an approximation for the similarity measure; research directions include subsequence matching and full sequence matching. The representation and similarity measure part of the time series analysis indexing is closely related; the research topic includes SAMs (SpatialAccess Methods) and TS-Tree [20].

E. Outlier detection

Anomalies or outliers, which are unexpectedly useful pieces of information present in raw data, can occasionally cause data mining methods to fail [2, 8, 25]. The characteristics of outlier data objects differ significantly from those of typical data objects. These characteristics might offer useful insight into some intriguing inherent characteristics. Finding patterns in data that are significantly different from the rest of the data using the right metrics is known as outlier detection. This pattern frequently contains useful information about the system's abnormal behavior that the data is describing. Distance-based algorithms use a geometric interpretation to determine the separations between objects in the data.

According to [37], there are four main types of outlier detection methods: statistical distribution-based outlier detection (Barnett and Lewis, 1994), distance-based outlier detection (Knorr and Ng, 1997; Knorr and Ng, 1998), density-based local outlier detection (Lee and Bang, 2013), and deviation-based outlier detection (Knorr and Ng, 2006), and density-based local outlier detection (Lee and Bang, 2013) (Knorr and Ng, 1997). To find outliers, rough sets-based algorithms introduce fuzzy rough sets or rough sets [8, 20, 54].

Additionally, mining patterns requires the use of more complex techniques, such as mining time-series data and streams of data, which are collections of temporal sequences. Streams and time series data in the Exabyte volume are produced by real-time systems, communication device networks, micro-sensor devices, telemetry devices, and online transactions. These data contain sequences of events obtained over repeated measurements of time with a very fast varying update rate [15]. Consequently, an algorithm must be capable of real-time, parallel, one-time scan, multilevel, and multidimensional stream processing and analysis.

IV. Recent work of Data Mining in IoT

Recent IoT applications that have gained popularity especially smart homes and ambient assistant living (AAL) [41, 44, 45]. These create a living and caring environment out of intelligent objects, making individual lives simpler, more technology friendly, even more comfortable, curable, and healthier. Embedded objects are capable of simultaneous interaction with other objects, people, internal servers, and the outside world. The ProPHeT decision-learning algorithm, created by Youngblood and Cook in 2007, learns strategy and manages the intelligent environment. To observe activities, they make use of the Episode Discovery sequential pattern mining technique, the Active LeZi algorithm, which forecasts upcoming events, and an automatically built hierarchical hidden Markov model, which learns an action strategy for the smart environment [45].

A system was developed for a smart environment that can identify and monitor user daily activities [35, 53]. The latter one used face recognition from visual data from the camera and audio data from a series of microphones with online Diarization capability. The former, on the other hand, used an unsupervised Discontinuous Variable-order Sequential Miner

to discover activities, then clustered them into groups and recognized them using a strengthened hidden Markov model. Naiive Bayes Classifier was used to identify activities. When there is a large volume of data, high observed value probabilities, and conditional independence of the features, this classifier performs well.

In order to get around this problem, Li et al. [68] highlighted the ineffective data collection methodologies used for activity recognition. As a result, for their research, they created a self-constrained, scalable, and energy-efficient custom WSN with a compact data format for episode mining.

Activity recognition and tracking in a smart environment have been used to make life easier for individuals and increase comfort [14]. This daily activity recognition and discovery can be used to predict behavior deviations in people who need to be medically monitored patients and receive emergency healthcare. Several work suggested a smart home with binary sensors to increase the autonomy of the patient under medical supervision [41, 44]. The strategy is divided into two sections: the first models user activities using Sequence Pattern Mining with Extended Finite Automation, and the second looks for behavior deviation using the residual method in case of a medical emergency.

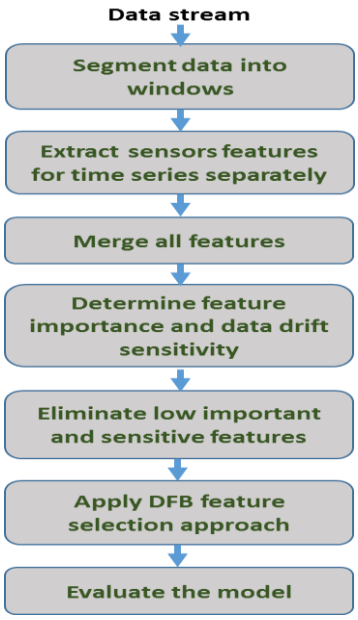

Fig. 3. Zdravevski et al. [19] approach for sensor data mining

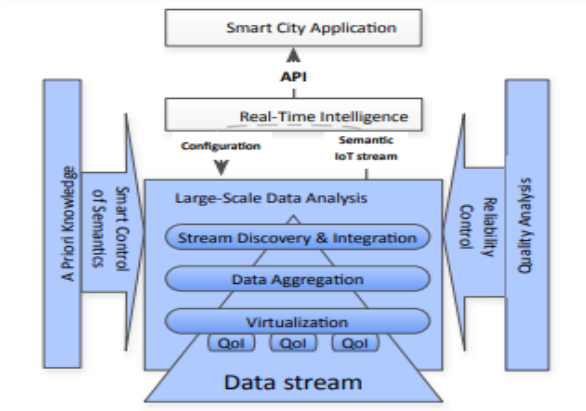

Fig. 4. Kolozali et al. [59] approach for memantine data pipeline-based real-time monitoring of the city's pulse

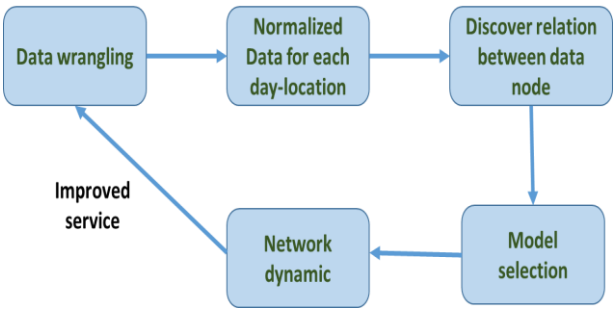

Fig. 5. Kotevska et al. [48] dynamic network for the resilience of data failure services

Zdravevski et al. [19] employed a variety of sensor data that was segmented with sliding windows, extracted time and frequency domain features using first deviation, Delta series, and Fast Fourier transformation, reduced features using Diversified Forward-Backward (DFB) Feature selection, and then generated classification models using machine learning algorithms like Logistic Regression, Extended Randomize Tree, and SVM with Gaussian kernel (Fig. 3).

Kolozali et al. [59] proposed a novel architecture that enables fast memantine data pipeline-based real-time monitoring of the city's pulse. The suggested architecture enables real-time data aggregation, semantic web framework, dynamic event handling, robust semantic convergence of data sources, and quality monitoring. The author evaluates their framework using real-time sensor observations published by the City of Aarhus on the open-source Open Data Aarhus platform (Fig. 4).

Kotevska et al. [48] developed a dynamic network model to increase the resilience of data failure services. The network model uses these trends to increase prediction accuracy in the absence of data by identifying statistically significant mutual time patterns through multivariant streams. These patterns are used by the network model. Due to the system's complexity, it frequently reacts to changes in the data stream by losing the ability to integrate new data flows (Fig. 5).

A blockchain-based cryptographic IoT data privacy-preserving SVM training method called Stable SVM is proposed [46]. A secure, dependable data-sharing network is created using blockchain technologies for multiple data providers to encrypt IoT data and then archive it on a distributed directory. They employ a homomorphic cryptosystem called Paillier that is intended to build strong building blocks like secure comparison, secure polyneme multiplication, and a secure SVM training algorithm that only needs two connections for one iteration and does not require a reliable third party (Fig. 6).

By fusing current Sanya tourism, All-for-One tourism, and smart cities, C. Xu et al. [12] developed Sanya tourism inspiration that examines the advantages and disadvantages of creating a Smart Tourism City in Sanya, and offers solutions to these problems. Understanding how to use the Internet and Big Data would become a Smart Cities of Sanya building strategy by creating a Wisdom Travel Framework using a mobile app.

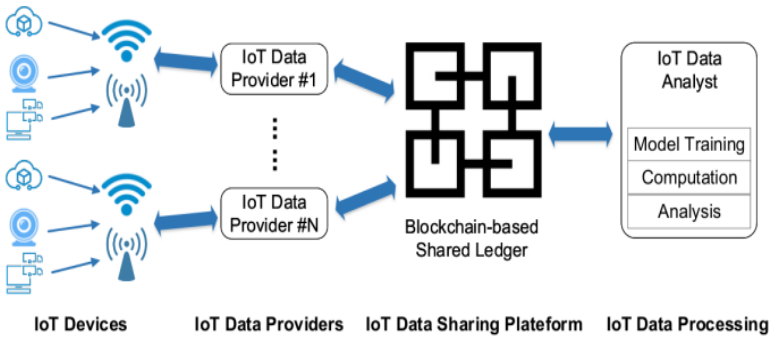

Fig. 6. A blockchain-based cryptographic IoT data privacy-preserving [46]

Others examined the study of big data urban digital management needs, mentioned the significant difficulties of early smart city development, and suggested a model and mitigation strategies for big data city development and activity. Amsterdam, for instance, has public spaces and sustainable transportation. Start with environmental protection and energy, then establish a sustainable environmental resource system and urban development strategy. Amsterdam, for instance, has public spaces and sustainable transportation [50].

An intelligent city architecture with Big Data Analytics (BDA) built in was proposed. The primary goal of the suggested Scheme is to implement adequate Big Data (BD) processing in order to increase the effectiveness of real-time decision-making. The review's findings offer insightful data for the development of the community while ensuring that the processing speed and throughput of the suggested framework are increased.

In classification, large-scale IoT, a set of classifiers can also implement the fusion of different classification techniques. Methods for classifying application issues include C4.5, a branch of CLS and ID3. It produces classifiers in the form of more understandable rules. When C5.0 came out, it replaced C4.5 with significantly greater efficiency, scalability, and performance by addressing drawbacks like high CPU and memory demands. The IoT environment of today is well-suited to classification models based on association rule analysis, support vector machines, and rule-based classification [9].

Researchers used classification models with frequent data mining techniques like Hidden Markov Models. The Markov Model can be used to create a more intelligent and responsive environment, where Naive Bayes, Gaussian naive Bayes, Bayesian belief network, Bayesian network, Artificial Neural Network, and Ensemble methods are applied to various sensor and actuator data in applications such as Biomedical, Environmental Prediction, smart building access controlling [33] and user activity recognition, improving Sensor network Efficiency, and so on [14].

EdgeMiningSim was proposed in [11], which is a simulation-driven methodology inspired by software engineering principles for enabling IoT Data Mining (Fig. 7). Edge Mining and Cloud Mining are Data Mining tasks aimed at IoT scenarios and performed using Cloud or Edge computing principles, respectively.

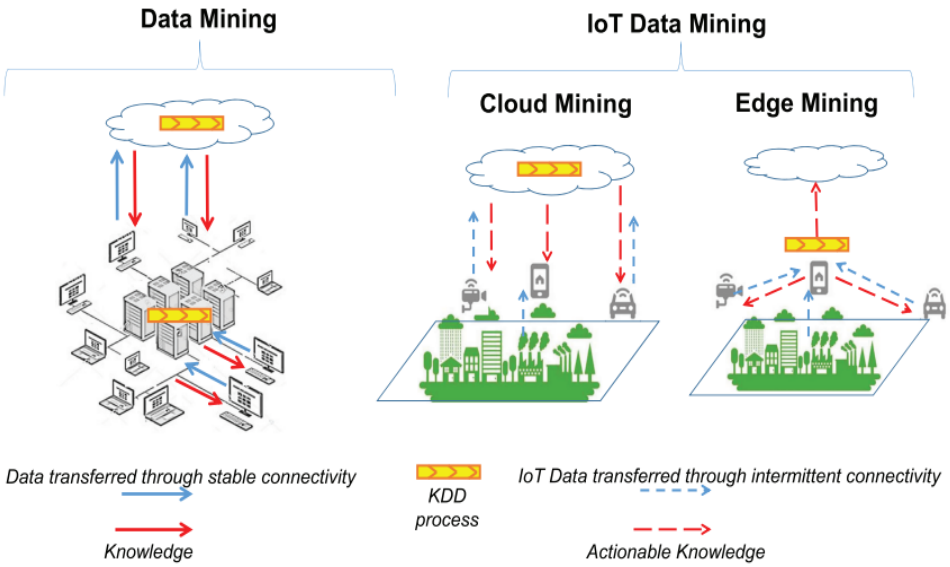

Fig. 7. A simulation-driven methodology using software engineering principles for enabling IoT Data Mining in cloud-edge computing environment [11]

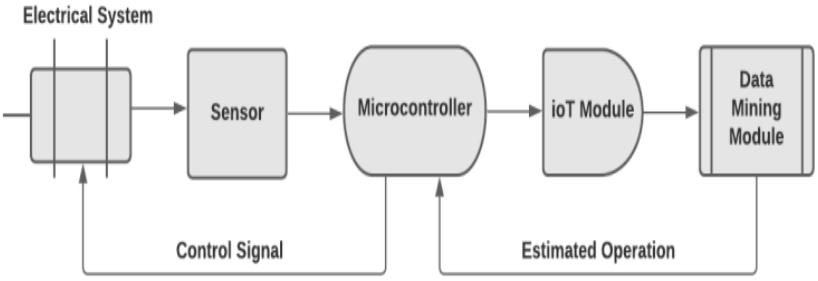

Fig. 8. Architecture of IoT-based data mining system for detecting the regularity of power generation and failure or defective regions in solar power systems [61]

Such a methodology motivates domain experts to reveal actionable knowledge, such as descriptive or predictive models for taking effective actions in a constrained and dynamic IoT scenario. A Smart Monitoring application is created as a case study to demonstrate the EdgeMiningSim approach and its benefits in effectively dealing with all of the multifaceted aspects that impact IoT Data Mining at the same time.

True-value-based differential privacy (TDP) is a novel privacy model proposed in [69] that applies traditional differential privacy to the "true value" unknown by the data owner or anonymizer but not to the "measured value" containing errors. Due to sensing errors, available data may differ from true data in many cases, particularly in the Internet of Things (IoT). Differential privacy, the de facto standard privacy metric, can be achieved by adding noise to a protected target value. There is no reason to add extra noise if the target value already contains errors. This paper discusses Based on TDP, a solution to reduce the amount of noise added by differential privacy techniques by about 20%.

The algorithm in [61] would assist human employers in detecting the regularity of power generation and failure or defective regions in solar power systems using IoT and data mining techniques (Fig. 8). This enables a quick action for fault rectification, resulting in increased generating station efficiency. Acres of land are required for the installation of large-scale power generating stations. Maintaining such a large area of power station is a difficult task for human employers. All electrical equipment requires calibration after a certain period of use to ensure its efficiency. These calibrated values are used in the data mining process in the proposed work. Thingsboard IoT platform is used in the work to monitor current and voltage variations. It will help the maintenance team monitor the virtual response of the solar power system that is connected. Simultaneously, the same data is sent to a local microcontroller, which compares the received value to the calibrated value.

The work in [31] presents an innovative approach to hazardous risk identification that relies on massive multisource data monitored in real time by the Internet of Things throughout the food supply chain. It addresses the risk identification and traceability system in food supply chains to identify risk levels and mine traceability rules. The proposed method's goal is to assist managers and operators in food enterprises in determining accurate food security risk levels in advance, as well as to provide regulatory authorities and consumers with potential rules for better decision-making, thereby maintaining the safety and sustainability of food product supply. The verification experiments show that the proposed method has the highest prediction accuracy.

Several work like [19, 53, 62, 68] classified numerous features that identify an individual's daily activities and automated their tasks to increase comfort and security using clustering as a core technology (Fig. 9). Cloud-based distributed clustering is more significant than centralized clustering in an environment like IoT and WSN because the data and the devices are highly distributed [66].

Cloud-based distributed clustering is more significant than centralized clustering in an environment like IoT and WSN because the data and the devices are highly distributed [30]. As a result, various processing techniques may be needed. The model proposed by Saives et al. [34] to find activity and identify behavior deviation. They use binary sensor event data to perform activity discovery, and they use Extended Finite Automation to cluster the activities into original final state models for further activity recognition. Sequential behavioral pattern discovery with frequent episode mining was provided by Li et al. [68] (FEM). FEM customized the DBSCAN clustering algorithm to mine both categorical and numerical data. Clustering can also be applied in IoT and WSN to make sensor network even more energy efficient, optimized [12] and to reduce transmission distance.

Deviation/outlier detection has been very helpful for IoT applications such as smart homes, smart agriculture, smart traffic and parking systems, healthcare, etc [38]. The e-health monitoring system prototype by Biswas and Misra (2015) uses biometric sensors and an Arduino UNO board to measure and gather individuals' essential health parameters. When a medical emergency arose, they used outlier detection mining to extract any anomalous information. In their cluster-based data analysis framework, Yu et al. (2017) proposed outlier detection with accuracy and redundant sensor data aggregation. To increase the efficiency of IoT-based systems, they used recursive principal component analysis (R-PCA). To reduce communication overhead and achieve much faster convergence, Zhang et al. [55] developed a decentralized approach based on Network anomaly detection. Several outlier detection techniques in IoT systems are discussed in details in [70]. It includes industrial, agriculture, financial, smart city, security and medical domains (Fig. 10). They categorized the source of outlier data to: error and noise, events, and malicious attack, which could be point, contextual or collective outliers.

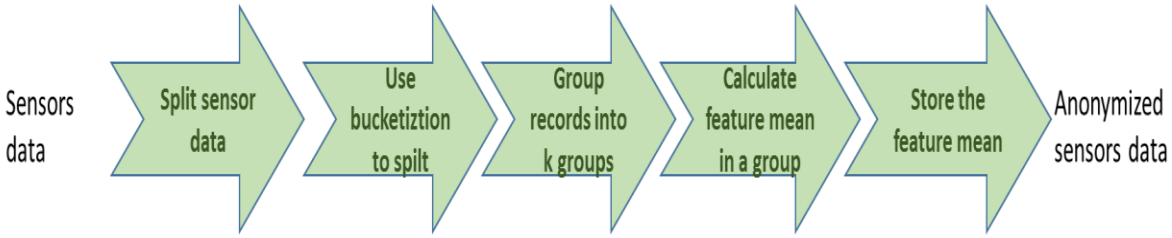

Fig. 9. The micro-aggregation approach to anonymize sensors data [62]

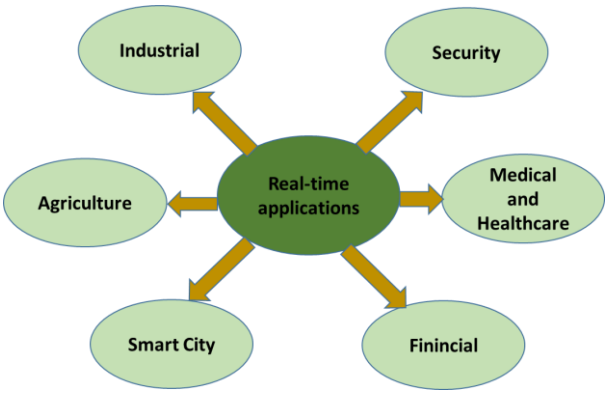

Fig. 10. Outlier detection techniques in IoT systems [70]

V. CONCLUSION

In this paper, the data mining approaches are discussed pointing to the most relevant approaches for the IoT field. The recent work for each type are highlighted and references for more detailed surveys are given. It is evadible that the field of data mining is growing to a large rate to enable understanding and managing the IoT systems in more intelligent and efficient manner utilizing the resources and reducing the costs.

VI. REFERENCES

[1] A. Broder, L. Garcia-Pueyo, V. Josifovski, S. Vassilvitskii, and S. Venkatesan, "Scalable K-Means by ranked retrieval," Proceedings of the 7th ACM International Conference on Web Search and Data Mining, pp. 233–242, February 2014.

[2] A. Elmogy, H. Rizk, and A. M. Sarhan, "OFCOD: On the Fly Clustering Based Outlier Detection Framework," Data, MDPI, vol. 6, no. 1, 2021.

[3] A. Iqbal, F. Ullah, H. Anwar, K. S. Kwak, M. Imran, W. Jamal, et al., "Interoperable Internet-of-Things platform for smart home system using Web-of-Objects and cloud," Sustainable Cities and Society, vol. 38, pp. 636-646, 2018.

[4] A. Kousis and C. Tjortjis, "Data Mining Algorithms for Smart Cities: A Bibliometric Analysis," Algorithms, vol. 14, 2021.

[5] A. Osman, "A Novel Big Data Analytics Framework for Smart Cities," Future Generation Computer Systems, vol. 91, pp. 620–633, 2019.

[6] A. Sarhan, "Fog Computing as Solution for IoT-Based Agricultural Applications," Smart Agricultural Services Using Deep Learning, Big Data, and IoT Book, IGI Global Publisher, pp. 46-68, 2021.

[7] A. Sarhan, "Cloud-Based IoT Platform: Challenges and Applied Solutions," Book chapter in Harnessing the Internet of Everything (IoE) for Accelerated Innovation Opportunities, IGI Global, pp. 116-147, 2019.

[8] A. Sarhan, A. Elmogy, and E. Mahmoud, "Enhancing Grid Local Outlier Factor Algorithm for better Outlier Detection," Artificial Intelligence and Machine Learning Journal (AIML), vol. 16, no. 1, pp. 13-21, March 2016.

[9] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, "Internet of Things in agriculture, recent advances and future challenges," Biosystems engineering, vol .164, pp. 31-48, 2017

[10] C.C. Aggarwal, C.C. Data Classification: Algorithms and Applications; CRC Press: Boca Raton, FL, USA, 2015.

[11] C. Savaglio and G. Fortino, "A Simulation-driven Methodology for IoT Data Mining Based on Edge Computing," ACM Transactions on Internet Technology, Vol. 21, Issue 2, pp 1–22, June 2021.

[12] C. Xu, X. Huang, J. Zhu, and K. Zhang, "Research on the construction of sanya smart tourism city based on internet and big data," in 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2018, pp. 125-128.

[13] Cisco. Cisco Annual Internet Report (2018–2023); Cisco: San Jose, CA, USA, 2020.

[14] D. Bouchabou, S.M. Nguyen, C. Lohr, B. LeDuc, and I. Kanellos, "A Survey of Human Activity Recognition in Smart Homes Based on IoT Sensors Algorithms: Taxonomies, Challenges, and Opportunities with Deep Learning," Sensors, vol. 21, 2021.

[15] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: Challenges‹ Issues, and Opportunities. In: Hong B., Meng X., Chen L., Winiwarter W., Song W. (eds) Database Systems for Advanced Applications, DASFAA. Lecture Notes in Computer Science, vol. 7827, pp. 1-15, 2013.

[16] D. Li, C. JianJun, and Y. Yuan, "Big data in smart cities," Scientific China Information Science, vol. 58, 108101, 2015.

[17] D. T. Larose, "k-nearest neighbor algorithm," Discovering Knowledge in Data: An Introduction to Data Mining, pp. 90–106‹ JohnWiley & Sons, 2005.

[18] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra‹ "Dimensionality reduction for fast similarity search in large time series databases," Knowledge and Information Systems, vol. 3, no. 3, pp. 263–286, 2001.

[19] E. Zdravevski, P. Lameski, V. Trajkovik, A. Kulakov, I. Chorbev, R. Goleva, N. Pombo, and N. Garcia, "Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering," IEEE Access, vol. 5, pp. 5262 – 5280, 2017.

[20] 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, Hindawi, Vol. 2015.

[21] F. Ganz, D. Puschmann, P. Barnaghi, and F. Carrez, "A practical evaluation of information processing and abstraction techniques for the internet of thing," IEEE Internet Things Journal, vol. 2, no. 4, pp. 340–354, 2015.

[22] F. Zheng and G. I.Webb, Tree Augmented Naive Bayes, Springer‹ Berlin, Germany, 2010.

[23] Google, 2017. What is big data. [online]. Available: https://cloud.google.com/whatis-big-data/.

[24] H. Habibzadeh, A. Boggio-Dandry, Z. Qin, T. Soyata, B. Kantarci, and H. Mouftah, "Soft Sensing in Smart Cities: Handling 3Vs Using Recommender Systems, Machine Intelligence, and Data Analytics," IEEE Communication Magazine, vol. 56, pp. 78–86, 2018.

[25] H. Rizk, S. Elgokhy, and A. Sarhan, "A hybrid outlier detection algorithm based on partitioning clustering and density measures," Proceedings of the 10th IEEE International Conference on Computer Engineering & Systems (ICCES), Faculty of Engineering, Ain Shams University, Cairo, Egypt, pp. 175-181, 21-23 Dec. 2015.

[26] H. S. Nagesh, S. Goil, and A. N. Choudhary, "Adaptive grids for clustering massive data sets," in Proceedings of the 1st SIAM International Conference on Data Mining (SDM '01), pp. 1–17، Chicago, Ill, USA, April 2001.

[27] H. Shariatmadari, S. Iraji, and R. Jantti, "From Machine-to-Machine Communications to Internet of Things: Enabling Communication Technologies," Internet of Things to Smart Cities: Enabling Technologies; CRC Press: Boca Raton, FL, USA, pp. 3–34, 2018.

[28] H. Shukur, S. Zeebaree, R. Zebari, D. Zeebaree, O. Ahmed, and A. Salih, "Cloud computing virtualization of resources allocation for distributed systems," Journal of Applied Science and Technology Trends, vol. 1, pp. 98-105, 2020.

[29] I. M. Ibrahim, "Task Scheduling Algorithms in Cloud Computing: A Review," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, pp. 1041-1053, 2021.

[30] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," Data Mining Knowledge Discovery, vol. 15, no. 1, pp. 55–86, 2007.

[31] J. Kong, C. Yang, J. Wang, X. Wang, M. Zuo, X. Jin, and S. Lin, "Deep-Stacking Network Approach by Multisource Data Mining for Hazardous Risk Identification in IoT-Based Intelligent Food Management Systems", Computational Intelligence and Neuroscience, vol. 2021, Article ID 1194565, 16 pages, 2021.

[32] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Identifying services for short-term load forecasting using data driven models in a Smart City platform," Sustainable Cities Society, vol. 28, 108–117, 2017.

[33] J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su, and B. Fang, "A survey on access control in the age of internet of things," IEEE Internet Things Journal vol. 7, no. 6, pp. 4682–4696, 2019.

[34] J. Saives, C. Pianon, and G. Faraut, "Activity discovery and detection of behavioral deviations of an inhabitant from binary sensors," IEEE Transactions on A Automation Science and Engineering, vol. 12, no. 4, pp. 1211–1224, 2015.

[35] J. Schmalenstroeer, and R. Haeb-Umbach, "Online Diarization of Streaming Audio-Visual Data for Smart Environments," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 5, pp. 845-856, 2010.

[36] J. Siryani, B. Tanju, and T.J. Eveleigh, "A Machine Learning Decision-Support System Improves the Internet of Things' Smart Meter Operations," IEEE Internet Things Journal, vol. 4, pp. 1056–1066, 2017.

[37] L. L. Tang, J. S. Pan, X. Guo, S. C. Chu, and J. F. Roddick, "A novel approach on behavior of sleepy lizards based on K-nearest neighbor algorithm," in Social Networks: A Framework of Computational Intelligence, vol. 526 of Studies in Computational Intelligence, pp. 287–311, Springer, Cham, Switzerland, 2014.

[38] L. M. Abdulrahman, S. R. Zeebaree, S. F. Kak, M. A. Sadeeq, A.-Z. Adel, B. W. Salim, et al., "A State of Art for Smart Gateways Issues and Modification," Asian Journal of Research in Computer Science, pp. 1-13, 2021.

[39] M. Abdulrazaq and A. Salih, "Combination of multi classification algorithms for intrusion detection system," Int. J. Sci. Eng. Res., vol. 6, pp. 1364-1371, 2015.

[40] M. Alaa, A. A. Zaidan, B. B. Zaidan, M. Talal, and M. L. M. Kiah, "A review of smart home applications based on Internet of Things," Journal of Network and Computer Applications, vol. 97, pp. 48-65, 2017.

[41] M.A.M. Alshewimy F.H. Elgendy, and A.M. Sarhan, "Fog-based Remote in-Home Health Monitoring Framework,", International Journal of Advanced Computer Science and Applications, The Science and Information (SAI) Organization, vol. 12, no. 6, pp. 247–254, 2021.

[42] M. Bermudez-Edo, P. Barnaghi, and K. Moessner, "Analysing Real World Data Streams with Spatio-temporal Correlations: Entropy vs. Pearson Correlation. Automatic Construction, vol. 88, pp. 87–100, 2018.

[43] M.H. Dunham, and S. Sridhar, " Data Mining: Introductory and Advanced Topics," Pearson Education: New Delhi, India, 2006.

[44] M.K. Hassan, A.I. El Desouky, S.M. Elghamrawy, and A.M Sarhan, "Intelligent hybrid remote patient-monitoring model with cloud-based framework for knowledge discovery," Computers & Electrical Engineering, Elsevier publisher, vol. 70, pp. 1034-1048, 2018.

[45] M. M. Sadeeq, N. M. Abdulkareem, S. R. Zeebaree, D. M. Ahmed, A. S. Sami, and R. R. Zebari, "IoT and Cloud Computing Issues, Challenges and Opportunities: A Review," Qubahan Academic Journal, vol. 1, pp. 1-7, 2021.

[46] M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, "Privacy-Preserving Support Vector Machine Training Over Blockchain-Based Encrypted IoT Data in Smart Cities," IEEE Internet Things Journal, vol. 6, pp. 7702–7712, 2019.

[47] M. W. Berry, and M. Browne, "Understanding Search Engines: Mathematical Modeling and Text Retrieval," SIAM, vol. 17, 2005.

[48] O. Kotevska, A. G. Kusne, D. V. Samarov, A. Lbath, and A. Battou, "Dynamic network model for smart city data-loss resilience case study: City-to-city network for crime analytics," IEEE Access, vol. 5, pp. 20524-20535, 2017.

[49] P. Anatharam, P. Barnaghi, K. Thirunarayan, and A. Sheth, "Extracting City Traffic Events from Social Streams," ACM Transactions on Intelligent System Technology, vol. 6, pp. 1–27, 2015.

[50] P. Bellini, M. Marazzini, N. Mitolo, M. Paolucci, D. Cenni, and P. Nesi, "Smart City Control Room Dashboards Exploiting Big Data Infrastructure (S)," in DMSVIVA, pp. 44-50, 2018.

[51] P. Katre, and A Thkare, "A Survey on Shortest path Algorithm for Road Network in Emergency Services," 2nd International Conference for Convergence in Technology (I2CT). pp. 393-396, 2017.

[52] P. Koukaras, C. Tjortjis, and D. Roussidis, "Social Media Types: Introducing a Data Driven Taxonomy," Computing, vol. 102, pp. 295–340, 2020.

[53] P. Rashidi, D. J. Cook, L. B. Holder and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment," IEEE Transactions in Knowledge Data Engineering, vol. 23, no. 4, pp. 527–539, 2011.

[54] P. Sunhare, R.R. Chowdhary, and M.K. Chattopadhyay, "Internet of things and data mining: An application oriented survey," Journal of King Saud University – Computer and Information Sciences, Vol. 34, pp. 3569–3590, 2022.

[55] Q. Chen, W. Wang, F. Wu, S. De, R. Wang, B. Zhang, and X. Huang, "A survey on an emerging area: deep learning for smart city data," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 3, no. 5, pp. 392-410, Oct. 2019

[56] S. C. A. Thomopoulos, D. K. Bougoulias, and C.-D. Wann, "Dignet: an unsupervised-learning clustering algorithm for clustering and data fusion," IEEE Transactions on Aerospace and Electronic Systems, vol. 31, no. 1, pp. 21–38, 1995.

[57] S. Guha, R. Rastogi, and K., Shim, "CURE: an efficient clustering algorithm for large databases," ACM SIGMOD Record, vol. 27, no. 2, pp. 73–84, 1998.

[58] S. Guha, R. Rastogi, and K. Shim, "ROCK: a robust clustering algorithm for categorical attributes," in Proceedings of the 15th International Conference on Data Engineering (ICD '99), pp. 512–521, March 1999.

[59] Ş. Kolozali, M. Bermudez-Edo, N. FarajiDavar, P. Barnaghi, F. Gao, M. I. Ali, et al., "Observing the pulse of a city: A smart city framework for real-time discovery, federation, and aggregation of data streams," IEEE Internet of Things Journal, vol. 6, pp. 2651-2668, 2018.

[60] S. Panda, "Security Issues and Challenges in Internet of Things," The Internet of Things: Breakthroughs in Research and Practice; IGI Global: Hersey, PA, USA, 2017; pp. 189–204.

[61] S. Shakya, "A Self Monitoring and Analyzing System for Solar Power Station using IoT and Data Mining Algorithms," Journal of Soft Computing Paradigm (JSCP), Vol. 3, No. 2, pp: 96-109, 2021.

[62] S. Samarah, M.G. Al Zamil, A.F. Aleroud, "An efficient activity recognition framework: toward Privacy-sensitive health data sensing," IEEE Special Section on Advances of Multisensory Services and Technologies for Healthcare in Smart Cities, vol. 5, pp. 3848-3859, 2017.
[63] T. Popović, N. Latinović, A. Pešić, Z. Zečević, B. Krstajić, S. and Djukanović, "Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study," Computers and electronics in agriculture, vol. 140, pp.255-265, 2017.
[64] T.S. Brisimi, T. Xu, T. Wang, W. Dai, W.G. Adams, and I.C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," Proc. IEEE, vol. 106, no. 4, pp. 690–707, 2018.
[65] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: a new data clustering algorithm and its applications," Data Mining and Knowledge Discovery, vol. 1, no. 2, pp. 141–182, 1997.
[66] V. Garcia-Font, C. Garrigues, and H. Rifa-Pous, "A Comparative Study of Anomaly Detection Techniques for Smart City Wireless Sensor Networks," Sensors vol. 16, 2016.
[67] W. Ejaz, and A. Anpalagan, "Internet of Things for Smart Cities: Technologies, Big Data and Security," Springer: Cham, Switzerland, 2019.
[68] Y. Li, X. Cheng, Y. Cao, and D. Wang, "Smart choice for the smart grid: narrowband internet of things (NB-IoT)," IEEE Intrnet of Things Journal, vol. 5, no. 3, pp. 1505–1515, 2017.
[69] Y. Sei and A. Ohsuga, "Private True Data Mining: Differential Privacy Featuring Errors to Manage Internet-of-Things Data," IEEE Access, Vol. 10, pp. 8738-8757, 2022.
[70] M.A. Samara, I. Bennis, A. Abouaissa, and P. Lorenz, "A Survey of Outlier Detection Techniques in IoT: Review and Classification," Journal of Sensors and. Actuator Network, vol. 11, no. 4, 2022.