ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

INTEGRATION OF CLOUD AND PARALLEL COMPUTATIONS WITH EFFICIENCY OF DATA MINING AND INTERNET OF THINGS BASED ON PRINCIPLES OF WEB TECHNOLOGY AND DISTRIBUTED SYSTEMS

FAIROZ Q. KAREEM

Energy Engineering Department Technical College of Engineering, Duhok Polytechnic University, Duhok, Iraq. Email: fairoz.kareem@dpu.edu.krd

NAJI ABDULLAH MAJEDKAN ABDULLAH

Business Administration Department Akre Technical College, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq. Email: naji.abdullah@dpu.edu.krd

ZRYAN NAJAT RASHID

Computer Network Department Technical College of Informatics, Sulaimani Polytechnic University, Sulaimani, Iraq. Email: zryan.rashid@spu.edu.iq

YOUSIF SUFYAN JGHEF

Computer Engineering Department College of Engineering, Knowledge University, Erbil, Iraq. Email: yousif.jghef@knu.edu.iq

TEBA MOHAMMED GHAZI SAMI

Computer Science Department Faculty of Science, University of Zakho, Duhok, Iraq. Email: teba.sami@uoz.edu.krd

NISREEN LUQMAN ABDULNABI

Business Administration Department Technical College of Administration, Duhok Polytechnic University, Duhok, Iraq. Email: nisreen.nabi@dpu.edu.krd

Abstract

Many analysis technologies are being incorporated into IoT in order to make it smarter; one of the most significant technologies is data mining. Data mining is the process of extracting hidden information from big data sets and identifying new, interesting, and potentially relevant patterns. Connecting everything on the planet through the internet may seem unattainable, but the "IoT" will drastically alter our lives in the near future by making many impossible attainable. In this paper the Internet of Things (IoT) is discussed first in this article. The features of data from IoT and data mining for IoT which is consists of classification, clustering and frequent patterns are then briefly reviewed. Finally, this article discussed and concluded the data mining applications with internet of things (IoT).

1. INTRODUCTION

Since was designed as a technological tool for smooth integration of conventional networks and networking items on the first day of the Internet of Things (IoT). In theory, IoT (Internet

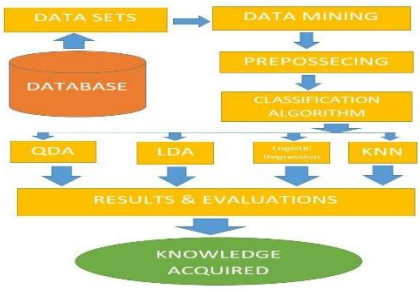
The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

of Things) connects everything in the globe. Things are assumed to be able to identify and, to a degree, communicate with one another and have the capacity to make judgments. An analysis carried out by Cisco predicts that the rise of the Internet of Things (IoT) as a service (IoT-as-AS) provider. The good news is that application development in the computer-to-to-machine world and the Internet of Things (IoT) will reach $44 billion in 2011, then to $290 billion by 2017. This means that researchers from a wide range of academic, industrial, and government sectors have invested in reinventing the internet because numerous intelligent devices, such as smart homes, smart transportation, and global supply chains, are available to be put in place that would significantly improve convenience [1].

Recently, numerous good surveys, each of which views the IoT with another viewpoint, have been given to explain what the IoT refers to: smartness, challenges , standards and applications. Atzori and his colleagues have offered among these studies the entire overview of IoT from three separate points of view, stuff, internet and semantics. A recent study inspired by showed that IoT was generically designed by five layers of architecture. From bottom to bottom, the five levels are the following: edge technology, middleware, gateway, Internet, and application. A number of more recently conducted studies, apart from outlining the infrastructures and the items in the area of IOT, have stressed that most of the items in IOT are intended to have intelligence, and thus dubbed (smart objects)[2] .

This is one of the biggest issues facing IoT today is how can we transform data into knowledge that is created or discovered to make our environment more pleasant for our clients? With knowledge discovery in databases and data mining, information in the IoT may be utilized to better the environment's function and efficiency. Data mining is therefore playing a crucial role in the IoT research and development process because of the number of investigations done. The findings in indicate that IoT-based solutions may be more intelligent by data mining algorithms [3]. In this paper, the remainder of the study is arranged as follows: section 1 begins with a quick introduction of the internet of things, followed by a discussion of why data mining tools should be used in the IoT. Section 2 discusses the characteristics of IoT data and data mining application with IoT which including classification, clustering, and frequent pattern mining, as well as the obstacles of acquiring, managing and analyzing IoT data. Section 3 is Related Work for researchers that worked on data mining with IoT in different fields. In the section 4 the possibilities for employing KDD for large scale data are discussed with IoT. Finally in section 5 draws conclusions and forecasts future development.

Figure 1: Data mining process



The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

2. BACKGROUND THEORY

2.1. Cloud Storage and Architecture

The phrase "cloud storage" refers to a specific paradigm for the storage of digital information in computers, in which the information is held in distinct data centers that are located in distant locations. The information is kept on a large number of servers, each of which might be housed in a separate data center. Cloud service providers are responsible for ensuring that the data is always available and can be accessed, as well as for the facility's safety, maintenance, and proper operation. In order to maintain the safety and integrity of user, corporate, or application data, businesses and individuals often rent or buy storage space from third-party suppliers. Connecting to cloud storage services is possible through colocated cloud computing services, web service application programming interfaces, or API-using applications. Examples of these types of applications include cloud desktop storage, a cloud storage gateway, and Web-based content management systems [3].

Cloud computing is commonly regarded as having been pioneered by Joseph Carl Robnett Licklider, who is credited with developing the concept via his work on ARPANET in the 1960s. 1983 was the year that CompuServe first started offering its consumer users free disk space for the purpose of data uploading and storing. AT&T established PersonaLink Services in 1994 as a web-based platform designed to facilitate interpersonal and commercial connections as well as corporate endeavors. They were the first companies to market their services using the term "the cloud," which referred to the electronic data storage facility that resembled a cloud and was used by the company. AWS S3, the cloud storage service offered by Amazon Web Applications, was first introduced in 2006. Since that time, it has been more popular as a storage provider for several well-known applications, such as SmugMug, Dropbox, and Pinterest. In 2005, the business cloud content management and file sharing service known as Box was first made available [4].

Since cloud storage is built on a similarly virtualized underlying architecture, it is identical to other aspects of cloud computing, including interfaces, almost immediate elasticity and scalability, multi-tenancy, and metered resources. This is because cloud computing as a whole is based on virtualization. Accessing and storing data in the cloud may be accomplished using either off-premises services (such as Amazon S3) or on-premises installations. Both are effective approaches (ViON Capacity Services). Cloud storage may take many forms, including object storage services hosted elsewhere, file storage, and block storage, just to name a few. The use of each of these unique types of cloud storage comes with its own set of advantages [5].

Object storage services can be hosted and deployed with cloud storage characteristics, and examples of such services include Amazon S3, Oracle Cloud Storage, and Microsoft Azure Storage; object storage software such as Open stack Swift; object storage systems such as EMC Atmos, EMC ECS, and Hitachi Content Platform; and distributed storage research projects such as Ocean Store [6] and VISION Cloud. File storage systems such as Amazon Elastic File System (EFS) and Qumulo Core are used by applications that need access to centralized or

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

distributed data storage. Use a Network Attached Storage (NAS) server to back up your data, whether it's a huge library, a development server, a video library, or even simply your home directory. NAS servers are becoming more affordable [6].

Block storage systems, such as Amazon Elastic Block Store (EBS), are essential to the operation of a significant number of commercial applications, such as databases, which need dependable access to massive volumes of data while experiencing as little lag as possible. This situation is comparable to that of storage area networks and direct attached storage (DAS), in certain respects (SAN).

Cloud storage, by its very definition, is: • Highly fault tolerant due to the redundancy and distribution of data; • Highly durable due to the creation of versioned copies; • Typically eventually consistent with regard to data replicas; • Made up of many distributed resources but still acts as one, either in a federated or a cooperative storage cloud architecture. Advantages Cloud storage has many advantages [7].

- Organizations are only required to pay for storage space that they actually use, which is often the average over a certain period of time such as a month, quarter, or year. The fact that cloud storage incurs operating expenses rather than upfront capital expenditures does not mean that it is less expensive. By putting their data on the cloud, businesses have the potential to cut their overall use of energy by as much as 70 percent, making them more friendly to the environment.
- Companies have the option of selecting off-site cloud storage, on-site cloud storage, or a hybrid combination of the two for their data storage needs. This choice can be made according to decision criteria such as continuity of operations (COOP), disaster recovery (DR), security (PII, HIPAA, SARBOX, IA/CND), and records retention laws, regulations, and policies. Initial direct cost savings potential is also a factor in this decision. Because they are already a part of the architecture of object storage, the additional effort, money, and technology that are necessary to ensure availability and security may potentially be avoided.
- The storage management obligations, including the purchase of additional storage space, will now be handled by the service provider. • This shift in responsibility will take effect immediately. Cloud storage provides users with a web service interface that enables them to have rapid access to a variety of resources and applications that are stored in the infrastructure of a different company.
- Using cloud storage, virtual machine images may be transferred between cloud and on-premises environments, or they can be imported from an on-premises environment into the cloud image library. Alternatively, virtual machine pictures can be exported from the cloud to an on-premises environment. The usage of cloud storage allows for the transfer of virtual machine images across the accounts of different users as well as between different data centers [7, 8].

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

Cloud storage offers protection against data loss in the case of a natural disaster since it often utilizes a large number of backup servers located in various parts of the globe. Users are able to access their cloud storage in the same manner as if it were a local drive thanks to WebDAV. It is possible for companies that have more than one location to utilize it as a centralized file server so that they may exchange files with one another. Concerns that may arise: An Exhaustive Analysis of Information Security Protection in the Cloud. When you keep your data away from the physical location of your business, you leave yourself up to the possibility of being attacked [9].

Because it is kept in a greater number of locations, distributed data has a greater risk of being physically corrupted. Because data is constantly being replicated and moved, the use of a cloud-based architecture, for example, significantly increases the possibility that illegal data recovery may occur. There are several instances of this, including the recycling of hard drives, the reuse of obsolete computers, and the reallocation of data storage space. The manner in which a consumer's data is replicated will be determined by both the service provider they use and the service level they choose. The use of encryption may secure critical data while still maintaining users' privacy. One of the available options for discarding sensitive information is known as crypto-shredding (on a disk) [10].

The likelihood of data being compromised (via improper methods, such as bribery or coercion) increases at an exponential rate whenever more people gain access to the information. The cloud storage provider will have a much larger team of technical staff with physical and electronic access to nearly all of the data at the entire facility or possibly the entire company. This is in contrast to the small team of administrators, network engineers, and technicians that might be responsible for the data of an individual company. If you compare this to the team that might be responsible for the data of an individual company, you'll notice that it's much smaller. When the user, rather than the service provider, is in possession of the decryption keys, access to the data by staff members of the service provider is limited. If users are going to be able to share numerous data sets in the cloud with a variety of users, then a large number of decryption keys will need to be sent to the users over encrypted channels before they can be used. The users will also need to be able to safely store and control these keys on their own devices. It is necessary to keep these keys in a secure place, which may result in additional costs. It is possible to get around this issue by using a cryptographic mechanism known as key aggregation. It expands the number of different networks that are available for information to go across. When it comes to connecting the data that is stored in the cloud, a wide area network (WAN) is required rather than just a local area network (LAN) or a storage area network (SAN) [11].

When data storage and networks are shared with a significant number of other users, it is possible for other customers to have access to the data that is stored there. It is possible that it is the result of hostile behavior; however, it is also possible that it is the consequence of human mistake, a malfunction, or a fault. This risk applies to any storage media, including cloud storage, and can't be completely eliminated. When data are encrypted, the risk of information being stolen while it is in transit is significantly reduced. The data that is being sent from your

ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

device to the cloud service may be encrypted while it is in transit to ensure its safety. Information that is retained with a service provider may be protected against unauthorized access by using encryption while it is "at rest." When data is encrypted in a system that is on-premises and connected to a cloud service, users have access to both varieties of encryption protection [12].

There are a number of various options available to you if you want to stay away from potential dangers. One option is to make use of a private cloud rather than a public cloud. This is one possible course of action. In addition, information may be imported in a secure manner, with the decryption key being kept locally. As a result, on-premise cloud storage gateways that include encryption options are often used in order to get access [13].

❖ **Longevity**

Businesses come and die, and the products and services that such businesses provide may change throughout the course of their existence. Due to the absence of assurances, investigation must be conducted extensively before outsourcing data storage. In the event that the company in issue declares bankruptcy or undergoes a significant change in its circumstances, any and all previously executed contracts may become null and invalid. The worst thing that might happen to a firm is that it may fail [14].

- Have your firm purchased by a larger one so that they may expand their operations and realign their priorities. Move because either your company was purchased by another company whose headquarters is situated in, or moved to, a country whose export compliance requirements make it hard for you to continue being present there, or because of a tragic catastrophe.
- Accessibility: The performance of outsourced storage is expected to be lower than that of local storage, however this difference is highly dependent on the customer's willingness to pay for wide area network (WAN) bandwidth.
- The reliability and availability of the service are closely correlated to the degree of measures taken by the service provider, as well as the accessibility of the wide area network. It is expected that both the hardware and the various algorithms would contribute to dependability. It is expected that a large number of storage locations are available [15].

❖ **Disparate Issues**

When saving sensitive information using a cloud storage service, you may be concerned about the protection of your data both while it is being maintained and while it is being delivered. This is because cloud storage services are not as secure as traditional hard drives. Cloud computing and storage may be difficult to utilize for some users, such as government agencies that are mandated by law to keep electronic data. These users may have difficulties. It's possible that both government security organizations and hackers might reap significant benefits from cloud storage. The cloud is an attractive target for hackers because it stores data belonging to such a large number of different people and companies. Websites that make it easy to share

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

data may also make it easier to steal content and violate intellectual property rights. The producers of material that is uploaded and shared on the network have taken legal action against the eBook storage provider Codex Cloud, just as they have done with its competitors Groove Shark and YouTube. When it comes to storing data either locally or abroad, maintaining regulatory compliance and the associated legal ramifications are a key concern. The resources needed to construct and maintain enormous data centers are contributing to a dramatic increase in the nation's overall production of energy. This has far more severe repercussions for the health of the ecosystem [16].

❖ **Storage in a Multi-Cloud Environment**

A storage system that makes use of both cloud storage and on-premises storage resources is referred to as "hybrid cloud storage," and the term "hybrid cloud storage" defines this kind of system. On-premises storage is often managed and protected by the firm that provides the storage, while public cloud storage is managed and protected by the storage provider. [33] Users are able to engage with the hybrid cloud in the same manner in which they would interact with a local storage system when using an on-premises cloud storage gateway. This kind of gateway may offer an interface for a file system or object storage. Thanks to its transparent data transfer to and from the cloud storage service, the cloud storage gateway enables low-latency access to the data via a local cache. This paves the way for low-latency access to the data [17].

Hybrid cloud storage provides businesses with the flexibility to utilize it either in combination with their already established in-house storage systems or as their primary storage system completely. Hybrid cloud storage often provides enterprises with more flexibility and scalability as compared to traditional on-premises storage systems [18].

Users of hybrid cloud storage have the ability to store hot data in local caches, from which it may be retrieved quickly, while cold data is held in faraway data centers, where it is protected by redundancy and is more difficult to access. There is a reduction in the amount of money spent on storage, as well as an improvement in both space and efficiency. Hybrid cloud storage provides organizations with higher redundancy and fault tolerance than traditional cloud storage does since it stores data in both on-premises and cloud-based systems [19].

2.2. Web Technology Principles

In order for the design of a website to be regarded effective, it must be capable of accomplishing the primary aim of the website, which is to educate visitors while also maintaining their interest in the website. Coherence, hues, colors, font, graphics, simplicity, utility, and functionality are some of the traits that a website that has been successfully designed would display in some combination [6, 20].

When it comes to developing the layout of a website, there are a number of key variables that need to be taken into account. Each of these factors plays a role in the overall appearance of the website. It should be the goal of any website to encourage users to take some kind of action and show that they have confidence in themselves. One of the most essential things that can be

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

done to guarantee that visitors have a great experience while they are on a website is to create a website that is both user-friendly and visually beautiful. This is one of the most important things that can be done (functionality). You can discover some useful suggestions for your next website project further down this page [21].

Your number one concern should be ensuring that people who use your website are happy with the experience they have there. If the purpose of each page is communicated in a way that is both clear and succinct, users are more inclined to engage with the content you provide. I would appreciate it if you could inform me as to the purpose of your website. Do the recommendations that you're providing here have any kind of instructional component? Do you provide content to your customers, such as sports reporting, or are you trying to sell them something in some other way? Websites are capable of performing a diverse array of tasks, but the following are some of the more common ones: identifying areas of speciality; creating a reputation; generating leads; selling products or services; and providing help once a transaction has been made [22].

Keeping things as straightforward and uncomplicated as you possibly can is the best way to enhance both the quality of the user experience and the functionality of your website. Several examples of various forms of minimalist design are shown in the following paragraphs. The term "navigation" refers to the method through which visitors to a website are able to learn about and explore various sections of that website. It is of the highest importance for a website to maintain its visitors' interest as they go through its many pages and to be able to do it effectively. If a website visitor is unable to find the information that they are looking for on a website that is difficult to navigate, the visitor will most likely leave the website. The navigation of the website should be user-friendly, consistent, and uncomplicated to utilize on each and every page of the website [23].

How to Make Sense of a Pattern That Looks like an F People who read content on websites often scan it in an F-shaped pattern as they read it. Eye-tracking equipment was used to conduct a research that found that individuals had a propensity to focus their attention on the left and top edges of the screen. The F-shaped pattern is a reflection of the Western nations' culture of reading (left to right and top to bottom). The eye of a reader who is scrolling down a page of a website that has been carefully developed will be guided in the appropriate direction [24].

Indicators of Visual History:

When referring to design, the term "hierarchy" refers to the order in which distinct graphical components are positioned in relation to one another. This order may be seen as a relationship between the elements. It is possible to accomplish this objective by the use of several design elements such as scale, color, image, contrast, typography, white space, texture, or style. A focal point draws the attention of site visitors to the aspects of a particular piece of content that are most noteworthy. This is one of the most important functions of visual hierarchy, as it directs the attention of site visitors to the aspects of a particular piece of material that are most noteworthy [25].

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

Not only will an amazing website have a beautiful design, but it will also include material that is really educational. When deciding whether or not to take action, such as making a purchase, website visitors may be swayed by the quality of the information presented on the site. Utilizing a style that is constructed on a grid may be of assistance to you in maintaining the organization and coherence of both your design and your content. The grid is a practical device that may be used in order to maintain the page's neatness and organization. The use of a grid-based method results in an experience that is both visually beautiful and delightful for the viewer since it organizes the material into columns and sections that are neat, consistent, and line up with one another while also seeming balanced [26].

If visitors to your website have to wait an excessive amount of time for the page to entirely load, you will almost certainly lose them as potential customers since your website loads slowly. If it takes a website more than three seconds to load, more over half of its users will move on to another choice instead of waiting for it to finish loading. It may be possible to dramatically reduce loading times on a website by optimizing the sizes of the photos used there [27].

Mobile-Friendly:

The number of individuals who access the internet via their mobile phones and other portable electronic devices continues to rise. When designing a website, it is crucial to keep in mind the requirement for a flexible style that is able to adapt itself to a variety of various screen sizes. This is something that must be kept in mind at all times. Building one's own website is a process that requires a lot of introspection on the part of the builder. Designers need to let go of their preconceived notions of how a successful website "should appear" in order to develop innovative and helpful compositions. This is necessary in order to generate original and usable compositions. Alternately, they should let the values of their brand, the audience for whom it is intended, and its unique selling proposition act as their "true north," allowing these aspects to guide every choice they make while crafting their designs. This would allow the brand's values, the audience for whom it is intended, and its USP to act as their "true north" [28].

There are a few tried-and-true web design principles that every designer has to bear in mind, despite the fact that design is all about breaking the "rules" and coming up with your own. These fundamentals have withstood the passage of time. You may design websites that consumers find engaging and are simple to browse by using these tried-and-true strategies, which can help you establish such websites. The next twenty of the most significant criteria for web design are listed here, and they need to serve as the guiding light for the work that you do moving forward. It would be unethical to downplay the significance that good typography plays in the process of designing websites because of how important it is. Picking the right fonts for your documents is a straightforward way to demonstrate the one-of-a-kind qualities of your company and capture the focus of your target audience. Nevertheless, there is more to it than simply how it seems at first glance. It is crucial that the fonts you choose to use in the design of your website are legible and easy to read. This is something you should keep in mind while making your choice. When everything is said and done, the user is less likely to be interested in what you have to say if they have to strain their eyes in order to read what you have written

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

and if the font you have chosen is difficult to read. The fonts Arial, Helvetica, Times New Roman, and Courier New are just a few examples of typefaces that are appropriate for use on the internet. They are readable at any size because of this, and they can be viewed on devices ranging from mobile phones to desktop PCs [29].

Matt Steel, who is both a writer and a designer, commissioned us to create a website for him in which text takes center stage on the page. He has employed a serif font, which has the benefit of being able to be read legibly in sizes ranging from very tiny to extremely big. He has done this by using the serif font. Do you want the typography in your next web design project to be able to make a statement all by itself and draw attention to itself? When you use Canva, you'll get a discount on the red geometric shapes that you purchase. When developing a header that is mostly consisting of text, the Website Ad template is an appropriate foundation to use as a starting point [30].

The fact that we humans are creatures of habit extends even to the kind of media that we choose to take in, which is why we tend to stick with what we know. The majority of us scan the material on a website using an F-shaped pattern, according to a research that was carried out by the Nielsen Norman Group using eye-tracking equipment. The information shown here leads us to believe that we should read the page in a clockwise direction, starting with the text that is bolded and the subheadings, and then reading the numbers, bullet points, and sidebars that are found along the left side of the page. The 'F' pattern is used in web design to avoid disrupting the viewer's line of sight by following the same route the viewer's eyes would naturally follow. This is accomplished by following the same path as the letter F. This eliminates any possibility of the viewer's line of sight being blocked in any way. This is an extremely important component for web sites that are primarily concerned with conversion, such as landing pages and sales pages. On the homepage of Big Commerce, which is a fantastic place to check it out, this can be seen in action. They have done an excellent job of accomplishing what it is that they want the user to achieve by using a layout that is shaped like the letter F. (in this example, check out the product tour). The title is the first thing that grabs your attention (in part because of the strong type and the colorful underlining), followed by the description of the functioning of the product, and then finally the supporting facts and the call to action at the end of the paragraph [31].

It is true that the 'F' pattern is often used for scanning with one's eyes; nevertheless, it is essential to keep in mind that this is not the only design that can be applied. There are other patterns that may be used. The usage of the 'Z' pattern is another essential factor that contributes to outstanding design. As the eye travels from left to right, a mental horizontal line is formed in the brain that runs from left to right as a result of this behavior. After then, it moves to the left side of the page, giving the impression that there is a diagonal across the page. At the very conclusion, it angles back to the right and, in the process, creates a second horizontal line. When compared to an "F" plan, then, under what conditions could a "Z" configuration be considered more appropriate? Landing pages that only include a limited amount of content function very well with the Z-pattern, which directs focus to the call to action (CTA) by using a horizontal scroll bar [32].

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

Even when everything around you appears to be going wrong, it is possible to find something positive to focus on during times like these. This is especially true in regard to the use of space on your website that is currently underutilized and is sometimes referred to as negative space. The gaps in your design that are not filled by any components are referred to as "white space" or "blank space," and they are regarded to be negative space (for example, the photos, text and icons). Even though it predates the history of art, this concept is critical to the development and success of any website. The last thing you want is for visitors to abandon your website because it is too difficult to browse due to excessive clutter, since this is the opposite of what you want to happen. The use of white space, on the other hand, makes it simpler to read, improves the overall experience for the user, and contributes to the elaboration of significant aspects of the design. In this particular illustration, the idea can be shown to be put into effect. Apple has never relied on marketing tricks that shift the attention away from the things it sells in order to boost its overall sales. One of the defining qualities of the design is the abundant use of white space, which serves the purpose of drawing exclusive focus to the iPhone X and is one of the design's distinguishing features. This may be seen. Do you plan to utilize the philosophy of "less is more" on your very own website and use it as your guiding principle? You may simply acquire a huge quantity of blank space with very little effort by adopting a design from Canva, such as the Pink Flower Wedding Events Website. This will allow you to save time [6, 33].

When you start working on the design of a website, you immediately realize how important it is to maintain uniformity throughout. It is not enough to just use the same fonts, colors, and symbols everywhere; it is essential to maintain consistency in the image of your brand. This requires ensuring that there is a constant amount of space between each element in each of your layouts by performing a check. This not only makes it more probable that customers would have trust in your firm, but it also gives the presentation of your website a more serious air. The website layout of Bookworm, a company that specialized in the provision of services for digital libraries, makes this point very clear. In spite of the fact that they employed a wide variety of text sizes and picture formats, the margin widths have been maintained consistently throughout the whole thing. In this way, harmony and order are brought to a piece of art that may have otherwise been disorganized. You can quickly guarantee that the text and paragraph spacing in your design are consistent with one another by using the in-built alignment tools that come with Canva. You can use these tools to make sure that your design is aligned properly [34].

It is a waste of time to visit a website that has navigation that is hard to comprehend since it is the same as being lost in a labyrinth without a guide. Because of this, the experience that people who use your website will have will be much more challenging and perplexing for them. If, on the other hand, the navigation of the website has been well considered and structured, utilizing the site could not only be easy but also enjoyable. This might be a pull-down menu, a sidebar, or even a navigation bar that is always present on the page. Accessibility, interoperability across devices, and the lack of unnecessary bells and whistles are the qualities that are regarded to be the most important ones [35, 36].

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

The atmosphere that is created by your website will be in large part determined by the color scheme that you choose for it. A color design that is heavy on dark browns and blacks may give off the impression that it is rustic and melancholy, while a color scheme that is heavy on pastels may give off the impression that it is contemporary and lively. You should make sure that the colors you pick all mix nicely together regardless of the style you are going for. This is something you should keep in mind while making your selections. Because of this, it is vital to choose colors that are comparable to one another in tone the majority of the time, despite the fact that it may also be pleasant to pick colors that oppose one another (for example, orange and teal) Choose colors that are either immediately across from one another on the color wheel or right next to one another; in either instance, the colors should be deemed to be complementary to one another. It is feasible to employ a wide variety of colors without sacrificing the aesthetic appeal of the overall design, as shown by the website that was designed for the Liebe Quark brand of yogurt. They were able to produce a strong visual impression by employing a color palette that contrasted warm and frigid tones with bright and dark ones in the header. This gave the header a more dynamic appearance. While the haphazard use of color contributes to a feeling of visual curiosity, the usage of a grid structure in the tiles helps to avoid the overall look from being too chaotic. Do you like a color scheme with less vibrant colors? It is easy to edit the Blue Fashion Blogger template that can be obtained on Canva so that it includes your own business logo and two brand colors that compliment one other [6, 37].

The best designers are conscious of the fact that making a website entails more than simply developing something that corresponds to one's personal aesthetic tastes, and they create websites with this knowledge in mind. If you want to build a website that stands out from the crowd and attracts the kind of visitors you're looking for, you need to have a comprehensive grasp of your target audience. Think about what it would be like to be a member of your target audience, and try to discover the most important problems, goals, and wants that they have by putting yourself in their shoes. It is essential to have this in mind when deciding on everything from fonts and colors to the text that appears on buttons and even the navigation of a website. Due to the fact that it is an internet business, this specific clothing company caters to the young female population as its target market. A color palette, iconography, and language that are powerful while yet being optimistic and uplifting were used in order to create the vivid and contemporary ambiance that is specifically designed to appeal to teenage females. You've decided that you want the design of your website to be identical like this one, right? You may utilize the Pink and Blue Rad Facebook Ad Template that Canva gives as a reliable foundation for the design that you come up with on your own. You can grab Canva here [38].

However, in spite of the significance of the part they play in the organization of the structure of a website, buttons are sometimes neglected throughout the design process. They have the ability to determine if the visitor stays on your website or closes the window and navigates to another one. The buttons on your website shouldn't have any type of modest aspect to them at all for whatever reason. To put it another way, they shouldn't be able to blend in with the background, and they should be simple to detect and touch at the same time. When we were discussing eye-scanning patterns and navigation, we spent some time discussing visual hierarchy. This is because eye-scanning patterns and navigation are intertwined. In any case, it

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

is a fundamental idea that is significant enough to be brought to the forefront on its own merits. Websites need not just to be imaginative and aesthetically pleasing, but they also need to make sense to the people who visit them. To put it another way, you need to proactively arrange your material in such a manner that it makes sense to the user, even if that sense is just on a subconscious level. This is something that you need to do in order to be successful. It is crucial to keep the broader picture in mind throughout the process of designing a website; doing so will help ensure success. After all, the aesthetic quality of your website as a whole is determined by the general visual composition of the website. Nevertheless, one should also pay attention to the specifics of the situation. The experience that a user has when using your website may be positively or negatively affected by apparently little design choices such as the micro interactions, font spacing, and other such details [39].

2.3. Integration of Multi-Technologies

- Learning in the twenty-first century is only possible with the assistance of technology resources that are effectively integrated and which are utilized by teachers who have received the required degree of training. Computers, mobile devices such as smartphones and tablets, digital cameras, social media platforms and networks, software programs, the Internet, and other forms of technology are all examples of tools that can be incorporated into the day-to-day operations of a classroom as well as a school as a whole. Other examples of technology that can be used in schools include 3D printers, virtual reality, and augmented reality. When technology is used in a way that is obvious and consistent, when it is easily accessible and immediately available for the work that needs to be done, when it helps to support the curricular objectives, and when it aids students in achieving those objectives, then technology integration has been successful. When technology is integrated into the classroom in a way that is effective and efficient, neither students nor instructors are even aware of when they are using it since it becomes natural to them. Students are more likely to be engaged in their work when digital technologies are included organically into the process of teaching and learning [1, 40].

- Before we can go on to addressing how to adapt our pedagogy or the role of the teacher in a classroom that is integrating technology, it is essential to first define exactly what is meant by the term "technology integration." Students are able to smoothly incorporate technology into their education and get a greater knowledge of subject matter when they not only utilize technology on a daily basis, but also have access to a selection of tools that match the work that needs to be done. However, the definition of technological integration may be altered depending on the kinds of technologies that are available, the extent to which individuals have access to technology, and the people who make use of the technology. If a classroom is just equipped with a single computer and an interactive whiteboard, for instance, the focus of education will most likely be on the teacher rather than on the pupils. Even an interactive whiteboard might be utilized to help you serve your students in a more effective manner. A willingness to learn and adjust is essential to the implementation of technology successfully. The field of technology is one in which progress is both steady and rapid. It is a process that never comes to a conclusion and requires ongoing research. When students are able to make educated judgments about which digital resources will best aid their research, analysis, and

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

presenting of results, technology integration in the classroom is at its most effective. In the same way that other resources in the classroom need to be freely accessible to all of the kids, so too must the technology [41].

- If they are used effectively, technology resources have the potential to greatly improve students' educational experiences. Access to recently published primary sources, means of obtaining and documenting information, strategies for working together with academics from all over the world, opportunities to demonstrate comprehension through the use of multimedia, relevant education, genuine evaluation, and training in presentation skills may be provided by tools such as these.

- Text editing via the use of computers and other technical devices is an example of editing technology, which is utilized across a range of topic areas to assist students in developing skills that are transferrable to other contexts. The majority of the time, the curriculum will decide how students will make use of technology in the classroom. Utilizing various forms of modern technology in order to enhance and complement the conventional learning environment. Incorporating technology into the classroom may help students learn in a number of different ways. One of these ways is by requiring students to do assignments on computers rather than the more conventional method of using pencil and paper. Another broader definition of technology integration is the process of integrating the many different SaaS (Software as a Service) applications, databases, and programs that are used by a school so that data can be shared in real-time across all systems on campus. This has the benefit of improving the quality of data and access for faculty and staff, which in turn benefits the education of students. Integration of technology into the learning process as a method of enhancing training across a variety of topic areas or in multidisciplinary settings, respectively. .. Students are better ready to incorporate technology into their education when they are able to make educated selections about which digital resources will best meet their requirements. In the same way that other resources in the classroom need to be freely accessible to all of the kids, so too must the technology. Rather of concentrating on the implementation of a single piece of technological equipment, each lesson or unit should be targeted toward the accomplishment of a particular learning goal. When technology is integrated into the standard curriculum, students have the potential to improve their self-confidence and get access to more advanced learning across a wider range of subject areas. However, the success or failure of integrating these technologies into educational programs is dependent on a variety of variables, one of which is the availability of sufficient infrastructure, which also requires constant maintenance and repair. In addition to the cost of the hardware and software, the operation and maintenance of technology integration in schools necessitates the expenditure of additional funds for things like electricity, Internet service providers, routers, modems, and people to manage the network. When technology is included into the normal educational curriculum, students have access to a broad range of cutting-edge learning tools that may be used for the study of a diverse range of topics. In today's world, when accountability, outcome-based education, and standardized testing are all prevalent, the usage of information and communication technologies (ICT) is often closely monitored and evaluated. There are several circumstances in which it could be challenging to implement new technology. According to the findings of several studies, it may be detrimental

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

to both instructional effectiveness and individual students' levels of productivity to allow many students to use the same computer or other technological resource. It has been shown that having peers work together on integrated technologies in pairs could promote a more cooperative attitude on social connections. [Citation needed] [Citation needed] There are a number of factors that exist outside of the realm of the technology itself that might influence whether or not the integration of new technologies is effective. Both students and instructors have an additional obstacle when it comes to gaining access to appropriate software for the integrated technology. One other issue that arises from the implementation of technological solutions is that there is insufficient long-term planning for the resources involved within the school districts that are putting these solutions into action. Technology in the classroom helps students build the core abilities they'll need to handle more difficult subjects. In addition to encouraging global awareness and diversity, this helps students become more open-minded. In order for technology to have an impact on the educational system, it must be made available to both students and teachers in a manner that is appropriate to their context, which is culturally responsive and meaningful to their educational practice, and which fosters quality instruction and student engagement. Only then will technology be able to have an impact [42].

- In order to compete effectively in today's global market, businesses need to analyze their operations, adapt to the changing needs of their customers, and continue to innovate in order to maintain their technical advantage. If a company hopes to stay competitive in their industry, it is imperative that it continuously updates both its procedures and its infrastructure. Before a change can be implemented, the reasons for making it must be stated, and those reasons must be in line with the larger goals of the business, which may include the pursuit of operational excellence, greater income, and other related goals. When introducing any form of process or technology adjustment, there are certain to be roadblocks at every turn. Before starting the process of integrating new technology, businesses have a responsibility to first evaluate the challenges that may be faced by their personnel as a result of the transition. If companies want to learn how the change will be beneficial to the people who will be touched by it, they should explore the reasons for their adoption of new technology. The following is a list of some of the most significant challenges that businesses face while attempting to implement new technologies:

- Failing to develop a comprehensive strategy for the digital transition: If there isn't a well-thought-out strategy that addresses the "What," "Why," "When," and "How" of the change, everyone who is engaged will experience a greater degree of frustration throughout the process of its implementation. It is essential to have a distinct and alluring vision of what the technology will become and what it will achieve. People who aren't informed on the ways in which new technologies might improve their lives and even shake up their routines for the better are more likely to be skeptical of such technologies when they first become available to the public. A thorough strategy program that outlines the benefits and downsides of the transition should involve the company, its personnel, its suppliers, and its customers as participants.

- The lightning-fast speed of technology advancement is becoming an increasingly important driver of the economy. The danger of falling behind the competition and losing out on new

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

opportunities is heightened for companies that do not stay current with the latest technical innovations and improvements. Because of this, you will need a staff that is both technically savvy and efficient, and one that is capable of redefining technology norms on the fly [43].

- A failure to communicate with and include all of the appropriate parties: It is essential, in order to accomplish early and speedy acceptance, to raise knowledge about the approaching integration. Transparency with regard to the new integration as well as the transformation method and its aims will go a long way toward achieving the goal of having the audience feel as if they are a part of the change. It could be good to explain the logical benefits as well as the financial advantages that the new service delivers to the corporation as well as the individual. Ensure that everyone is aware of the ways in which the new equipment improves upon the previous one. Because users place a high priority on functionality and simplicity of use, it is important to take into consideration the demands and preferences of these users.

- A deficiency in "Change Champions": It is essential, at an early stage in the implementation process, to form a group of individuals who will advocate for the new technology and inspire excitement among the employees. However, there is a deficiency in this area. After the announcement has been made, the group may meet privately with the individuals who will be impacted to address their concerns, answer their questions, and explain the benefits that will result from the change.

- A failure to identify, manage, train, and upskill persons who may be impacted by new technologies. a failure to identify, manage, train, and upskill individuals who may be affected by emerging technologies. It is advised to run comparative pilots and encourage them to complete trials and gather feedback in order to impact the sort, style, and intensity of training that is required. This is because persons who are not tech savvy may see training as a problem.

- Uncertainty: Prior to integrating or implementing any new technology, decision-makers need to spend time investigating and identifying potential risks that could render the new technology inoperable or useless. This research and identification must take place before any new technology can be integrated or implemented. In order to avoid this, it will be necessary to conduct a proof-of-concept pilot project, during which technical feasibility will be evaluated.

- Failing to have a backup plan; new technologies always have the possibility of introducing new issues and security flaws. It is vital to have a well-thought-out strategy for the introduction of any new technology, including actions to take care of difficulties that were not anticipated.

- In conclusion, in order to ensure the delivery of the new technology and its widespread implementation, acceptable expectations need to be established, the impact ought to be grasped, and obstacles and dangers ought to be investigated and dealt with in an appropriate manner [44].

2.4. Distributed and Parallel Systems Computing Principles

- A system is said to be distributed when its component elements are dispersed among several computers that are all linked together in some kind of network. These computers are able to communicate with one another and function together by sending and receiving messages. The

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

field of distributed computing is primarily concerned with the investigation of distributed systems. In a distributed system, each node is responsible for communicating with all of the other nodes in the system so that the nodes may collaborate on a single endeavor. In distributed systems, one of the most important challenges is maintaining concurrency amongst the system's components. Another challenge is managing the failure of individual components independently. In the case that only one component of a system fails, the functionality of the system as a whole will not be compromised in any way. Examples of distributed systems include a wide variety of various types of systems, such as those that are based on service-oriented architectures (SOAs), massively multiplayer online games (MMOs), and peer-to-peer applications. The process of developing computer programs that are capable of operating in a networked yet decentralized setting is referred to as "distributed programming." There are a variety of various ways that message transmission methods may be implemented. These include employing RPC-like connections, message queues, or even simply plain old HTTP. Distributed computing may also be defined as the practice of using several computer systems to the completion of a single job. Distributed computing enables several computers to collaborate on finding a solution to a problem by breaking down the issue into smaller, more manageable parts. This allows the computers to communicate with one another, share information, and work together to find a solution [45].

- Computer networks in which individual computers were physically dispersed throughout some geographical region provided the initial inspiration for the use of the word "distributed" in words such as "distributed system," "distributed programming," and "distributed algorithm." [C]omputer networks in which individual computers were physically dispersed throughout some geographical region provided the initial inspiration for the use of the word "distributed." The phrase has a somewhat larger range of application these days, and it is even possible to use it to represent independently executing programs that share the same physical system and exchange messages with one another. In spite of the fact that there is no one definition of a distributed system that is recognized by everyone, the following features are often used as stand-ins:

- Individual computers, also known as nodes, are capable of operating independently and storing data in their own respective memories. Entities (such as these) communicate with one another by sending and receiving messages.

- In a distributed system, a collection of individual computers may seem to a user as a single entity if they are all working toward the achievement of the same goal, such as the resolution of a complex computing problem. On the other hand, each computer may serve a single user who has very particular needs, and the distributed system may be in place to either make it easier for users to coordinate their usage of shared resources or to provide users access to channels of communication. Additional traits that are often seen in distributed systems include the following:

- It is essential that individual computer issues do not bring the whole system to a halt.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

• Each computer only has access to a limited and fragmented perspective of the system; the structure of the system, including its network topology, network latency, and number of computers, is unknown in advance; the system may be composed of a variety of computers and network links; and the system may undergo transformations while a distributed program is being carried out. It's likely that each computer can only see a portion of the input data that's been sent in [46].

• A distributed system is a group of computers that collaborate to complete a certain job by communicating with one another through a network. There is no meaningful distinction to be made between the ideas of "concurrent computing," "parallel computing," and "distributed computing," which all have many characteristics in common. Because the processors in a typical distributed system work in parallel with one another, a system may be described as "parallel" while still being described as "distributed." It is feasible to conceive of distributed computing as a kind of parallel computing that has loosely linked components, and vice versa, with parallel computing being a particular sort of distributed computing that has tightly coupled components. Notwithstanding this, concurrent systems may be classified as "parallel" or "distributed" based on whether or not they exhibit the following characteristics:

• When doing distributed computing, each processor has its own private memory; but, when performing parallel computing, all processors may share a common memory so that data may be exchanged between them (distributed memory). In order to make communication as easy as possible, messages are sent from one processor to the next.

• The research that was done in the 1960s on the designs of operating systems prepared the way for the use of several processes that collaborated with one another and sent messages to one another.

• [24] The first distributed systems to see widespread usage were local-area networks like Ethernet and similar technologies. The decade of the 1970s saw its inception. Email was first created for use on the ARPANET network in the early 1970s, making it one of the earliest forms of communication used on the Internet. It is possible that electronic mail was the first instance of a large-scale distributed application [26], and it went on to become the most used application of the ARPANET. Usenet and Fido Net, both of which were established in the 1980s, were among the first worldwide computer networks. These networks, together with ARPANET, enabled distributed discussion systems (and its successor, the global Internet). The study of distributed computing arose as a separate topic within the science of computing in the latter half of the 1970s and the early 1980s. [C]omputer scientists now refer to this research as "distributed computing." In 1982, the first Symposium on Principles of Distributed Computing (PODC) was held, and in 1985, the International Workshop on Distributed Algorithms on Graphs was held in Ottawa. The International Workshop on Distributed Algorithms on Graphs was the forerunner to the International Symposium on Distributed Computing (DISC), which was held in subsequent years [47].

• The practice of distributed computing calls for the use of a wide variety of computer hardware and software platforms. A network, whether it be printed onto a circuit board or comprised of

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

loosely linked devices and cables, is necessary on a basic level in order to connect a number of central processing units (CPUs). It is necessary to have a communication system in order to connect up the processes that are running on those CPUs in order to move things to the next level. In a distributed context, coding may be categorized as either loosely coupled or tightly coupled, or as client-server, three-tier, n-tier, or peer-to-peer architecture. Other possible classifications include n-tier and peer-to-peer [48].

• Systems that allow "smart clients" to connect with a central server in order to get information and then provide that information to end users in an appropriate manner are referred to as "client-server" systems. When a client makes a change that should be stored permanently, they must click a button on their device in order to transmit the information back to the server where it will be processed.

• One example of this would be three-tier architectures, which make it possible to have stateless clients by outsourcing the processing of those clients to a different layer of the network. Because of this, releasing brand new software is a considerably simpler process. On the internet, multi-level applications are the standard.

• N-tier designs, in which web applications often forward requests to several different corporate services this kind of software may very well be held directly responsible for the widespread adoption of application servers.

• Architectures known as peer-to-peer are those in which no specialized computers are employed to deliver services or manage the resources of a network. In its place, each and every work is equitably spread over all of the computers. Peers may operate in either a client or a server capacity depending on the situation. This kind of architecture is used by both the BitTorrent network and the Bitcoin network [49].

• An additional basic component of a distributed computing system is the mechanism that facilitates the exchange of information and the coordination of actions among processes that are currently active. Processes may communicate with one another via the use of message transmission protocols and can have a master/slave relationship with one another. When carrying out distributed computing, another alternative is to employ what is known as a "database-centric" architecture, which depends on a single data store rather than direct communication amongst the units that make up the system. The relational processing analytics that are supplied by the schematic structure of database-centric architecture are very helpful for real-time situations. Because of this, it is possible to make advantage of distributed computing both inside and outside the confines of a traditional networked database [50].

• There are a number of potential advantages to using distributed computing and distributed systems. There are certain applications that, because to the very nature of what they do, need the use of a communication network that links numerous computers together. One example of this is when data is created in one region but is required in another region. Even though it is theoretically possible to utilize only one computer, many times it is more effective to employ a distributed system instead because of the increased efficiency it provides. Distributed systems have several advantages over non-distributed ones; for example, they can: • provide

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

significantly larger storage and memory, quicker computation, and higher bandwidth than a single computer; • give greater dependability than a non-distributed system, since there is no single point of failure; • give greater dependability than a non-distributed system, because there is no single point of failure. A centralized system with a single processor may be more difficult to operate and expand, while a distributed system with several processors may be simpler.

• When compared to the use of a single high-end computer, the use of a cluster consisting of a number of low-end computers to attain the required level of performance may prove to be more cost-effective [1, 6, 51].

• The majority of the chores that we would want computers to do for us are of the "question and answer" kind. This means that we want to ask the computer a question, and it should provide us with an answer. These kinds of endeavors are referred to as computational issues in theoretical computer science. A computational problem may be stated in its most strict form as having instances and solutions for each individual instance. Instances are questions that we may ask, and solutions are the desirable replies that we might provide to these inquiries.

• Computability theory is a topic that is of special interest to theorists working in theoretical computer science since it focuses on computational challenges that can be successfully addressed with a computer (computational complexity theory). It is a widely held concept that in order to solve a problem, one must first be able to teach a computer to always return the correct answer to a given question before using the computer to solve the problem. This sort of procedure may be implemented on a general-purpose computer by writing it as a program that accepts an example of the problem as input, does some calculations, and then spits out the solution as output. Formalities such as random-access machines and universal Turing machines may be used in order to serve as abstract models of a sequential general-purpose computer that is responsible for performing such an algorithm.

• The study of concurrent and distributed computing seeks to answer the same kinds of questions for systems that consist of numerous computers as it does for a single computer that runs a network of interacting processes. These questions include the following: which computational tasks can be performed in such a network, and how efficiently can they be performed? In the context of a concurrent or distributed system, however, it is not at all apparent what is meant by the phrase "solving a problem." For example, what is the function of the algorithm designer, and how can we imagine a system that is equivalent to a sequential general-purpose computer?

• Even while many of the considerations that will be covered below are equally pertinent to concurrent processes that are working on a single computer, the discussion that will follow will focus on the situation in which there are numerous computers [52].

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

The following are three frequent points of view:

- Implementing parallel algorithms with a shared-memory architecture
- All of the central processing units have access to a single memory that is referred to as shared memory. The developer of the algorithm chooses which version of the program will be executed on each CPU.
- Parallel random-access machines, often known as PRAMs, are used as a model for theoretical purposes.
- [40] However, in the typical PRAM design, synchronous access to the shared memory is seen as a given rather than a need.
- Shared-memory programs may be ported over to distributed systems if the operating system is capable of isolating the connection between nodes and effectively unifying the memory across all of the distinct computer configurations.
- Asynchronous shared memory is a more accurate representation of how real-world multiprocessor computers behave because it takes into account the execution of machine instructions such as compare-and-swap. This makes asynchronous shared memory the most common type of memory used in modern computers (CAS). There has been a significant amount of study done on this idea, and the scholarly literature may provide a synopsis of some of this work.
- Parallel algorithms that communicate with one another using a message-passing model.
- The creator of the algorithm is the one who chooses the design of the network as well as the code that is executed on each computer.
- Both Boolean circuits and neural networks are used as models in this process. Each gate in a Boolean circuit is analogous to a minicomputer that is responsible for carrying out a fundamental procedure. It is possible to think of each comparator in a sorting network as a separate computer.
- Models of distributed computing that are dependent on the transmission of messages
- The developer of the algorithm just has to choose the application to use for the program. The identical piece of software is installed on each and every system. It does not matter how the network is organized; the system must still operate correctly.
- An increasingly common conceptualization is a network in which each node stands for a distinct finite-state computer [53].

2.5. Systems Efficiency Complexity Measures

Consider the computational difficulty of coloring the graph G given its vertices and edges as an example.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

The following are some possible tactics that might be used in a variety of different fields:

The compilation of algorithms at a single place

A computer is given Graph G in the form of a string, which is a representation of the graph's encoding. A string representation of the coloring that the computer has discovered for the graph is generated by the computer.

Algorithm parallelism

To reiterate, a string representation is used to depict the graph G. On the other hand, many computers are able to access the same string at the same time. It's likely that different computers will color certain parts of the graph in a different way. The use of a large number of computers working together in parallel in order to achieve high-performance computing is a fundamental problem [45].

Processing that is distributed in Algorithms

The graph denoted by the letter G may serve as a representation of the architecture of a computer network. The vertices (nodes) of G are each represented by a computer, and the edges (edges) of G are each represented by a communication link. Because each computer initially knows information about just its near neighbors in the graph G, they are need to interact with one another in order to acquire further information about the structure of G. During the printing process, each machine must use a color that is entirely distinct to itself. The management of the proper operation of a totally decentralized system is the fundamental objective. Even if parallel algorithms have a more laser-like focus, there is still a significant amount of overlap with distributed algorithms. The Cole-Vishkin technique, which colors graphs, was first described as a parallel algorithm; however, it may also be utilized directly as a distributed algorithm by using the same strategy. Additionally, a parallel technique may be implemented in either a parallel system (with shared memory) or a distributed system. Both of these configurations are possible (using message passing). The traditional distinction between parallel and distributed algorithms (choose the best network vs. use any old network) and parallel and distributed systems does not correspond. Choose the best network vs. use any old network (shared memory vs. message passing) [54].

Complexity is being analyzed here.

Parallel algorithms have the ability to make use of a variety of resources, including time, space, and the number of computers that are available. In point of fact, there is often a trade-off between the amount of time the program is allowed to run and the number of processors: the problem may be resolved more quickly with more computers working in parallel (see speedup) If a decision problem can be solved in polylogarithmic time using a polynomial number of processors, then the problem belongs to the class NC of decision problems. Given that PRAM machines are able to successfully imitate Boolean circuits, it is possible to define the NC class using either the PRAM formalism or Boolean circuits. Both approaches are sufficient [46].

When analyzing distributed algorithms, the attention is often placed more on the communication activities than on the computing ones. A synchronous system, which is

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

possibly the simplest kind of distributed computing, is one in which all nodes operate in lockstep with one another. This framework is also known by its more common moniker, the LOCAL model. During each cycle of communication, all of the nodes concurrently carry out three tasks: (1) they receive the most recent messages from their neighbors; (2) they carry out some arbitrary local computation; and (3) they send new messages to their neighbors. When evaluating the complexity of such systems, one method that may be used is to count the number of iterations of synchronous communication that are required to complete the task [55].

In a manner similar to how the length of a network has been shown to have a correlation with this complexity measure. Take into consideration that the network has a diameter of D. On the one hand, when using a synchronous distributed system, it is simple to solve any computable issue in about 2D communication rounds. This is accomplished as follows: gather all of the data in one location (D rounds), solve the problem, and then distribute the results to the nodes (D rounds). However, in the event that the running time of the algorithm is shorter than D communication rounds, the nodes of the network will be required to produce their output without having access to information about further-flung parts of the network. To put it another way, each node must behave in a manner that is consistent with the actions of the other nodes in the network while just using the data that is available from its local D-neighborhood. The question of whether or not a certain issue can be solved by one of the several distributed algorithms that are known to exist in a running time that is noticeably less than D rounds is one of the most urgent difficulties in this area of study. [49] In this context, efficient algorithms are often those that do their job in a time that is polylogarithmic in relation to the size of the network they are working on. Another common measurement is the total number of bits that are sent over a network (cf. communication complexity). It is common practice to utilize the CONGEST (B) model to express the components of this concept. This model was developed in the same manner as the LOCAL model, with the exception that individual messages may only contain B bits [56] .

In addition to this, there are further problems.

A user submits a question, the question is processed by a computer (or distributed system), and then the system is shut off after the question has been answered. This is the traditional approach to solving computational issues. In some circumstances, such as the one involving the dining philosophers and the associated problem of mutual exclusion, it is essential to ensure that the system is operational at all times. In order to solve these problems, the distributed system has to be able to coordinate the use of shared resources in a dependable manner in order to prevent unfavorable outcomes such as deadlocks and disagreements.

In addition, distributed computing comes with its own unique set of fundamental challenges, such as those revolving around the concept of fault tolerance. Concerns pertaining to consensus, Byzantine fault tolerance, and self-stabilization are only a few examples of related challenges. The asynchronous behavior of distributed systems is another topic that receives a lot of attention from researchers.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

Synchronizers enable for the execution of synchronous algorithms inside asynchronous systems, logical clocks offer a causal happened-before ordering of events, and clock synchronization methods provide globally consistent physical time stamps [57].

The goal of coordinator elections, also known as leader elections, is to pick a single process that will be responsible for coordinating the activities of several computers in order to accomplish a certain task (nodes). Before a job is begun, all of the nodes in the network are either unaware of which node will serve as the "coordinator" (or leader) of the task, or they are unable to communicate with the node that is already serving in that capacity. However, after a procedure for the election of a task coordinator has been carried out, every node in the network will reach a consensus on a single node to serve in the role of task coordinator [58].

The connected devices in the network engage in communication in order to ascertain which device will play the function of "coordinator." To achieve this goal, they need to devise a strategy that will throw off the natural equilibrium that exists between them. If each node, for example, has an identity that is separate from but yet comparable to the identities of the other nodes, then the nodes may use this comparison to choose which node will act as the coordinator of the network. LeLann is mostly credited for formalizing the notion of this issue as a mechanism to create a replacement token in a token ring system when a token has been lost. This issue was a way to solve a problem that had been plaguing the system for some time [59].

The algorithms that are used to choose a new coordinator have been improved so that both time and data may be saved. The generic undirected graph algorithm that was proposed by Gallager, Humblet, and Spira, and which was awarded the Dijkstra Prize as the most outstanding publication in distributed computing, has had a significant influence on the design of distributed algorithms in general. The generic undirected graph algorithm was awarded the Dijkstra Prize as the most outstanding publication in distributed computing. Some of the network graphs for which alternative approaches have been offered are undirected rings, unidirectional rings, complete graphs, grids, and directed Euler graphs. Other network graphs include grids. The researchers Korach, Kutten, and Moran came up with an idea for a universal method that would remove the graph family issue from the design of the coordination election algorithm. In distributed systems, coordinators are used to assist in the completion of activities that need some degree of coordination. The problem of electing a coordinator involves selecting a winner from a field of candidates who are each operating on a different node inside a distributed system. There are a few different algorithms that may be used to choose a central leader [60].

Decentralized systems have the following characteristics:

Up to this point, the primary focus has been placed on the development of a decentralized system to handle a particular problem. One field of investigation that is connected to this topic is the study of the properties of a particular distributed system.

When we are given a computer program and asked to decide whether or not it halts or continues endlessly, we encounter a difficulty that is analogous to the one that may be found in the study of centralized computing. This problem is known as the halting issue. It is intrinsically at least

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

as difficult to comprehend the behavior of a computer network as it is to understand the behavior of a single computer, and the stopping question is undecidable in the general scenario.

Despite this, there are several fascinating edge cases that may be resolved. In particular, it is possible to reason about the activities carried out by a collection of linked finite-state machines. Consider, as an example, the challenge of figuring out whether or not a certain network of interacting (asynchronous and non-deterministic) finite-state machines may be approaching a stalemate. Even if it is theoretically feasible to solve this problem by utilizing a computer, it is very improbable that an effective solution—either centralized, parallel, or distributed—exists, which would make the problem PSPACE-complete.

2.6. Data from IoT

Technically, anything connected to the internet of things might result in a data avalanche containing a variety of useful data. For some years, the data and findings have presented more difficulties when it comes to discovery and use, especially in recent years, especially as information has become increasingly complex. To aid with differentiation of IoT, a taxonomy [61] divides data that refers to things into two classes: “data about things” (such as status, location, and identification) and “data created by things. Normally, the former includes data that may be used to optimize systems, structures, infrastructures, and services, while the latter includes data that is user-generated and available for use in such systems [62].

The new form of data gathered by sensor or RFID has been characterized as “big data.” These features allow us to immediately conclude that big data is no longer just a business; it's here to stay. Since studies throughout the world are estimating different quantities of data are created every year, it is projected that the overall amount of data has increased by a zettabyte in the last decade. Today's data analysis techniques are not sufficient to manage the complexity of data from the Internet of Things (IoT) Putting several zettabytes into a single storage system will always be tough. Next, we need to address the issue that most IoT data is large and difficult to handle by today's technologies. The rate limiting factor will be moved from sensors, followed by their ability to transmit data, followed by data storage capacity. The analysis also demonstrates that user interface design and development will have to be re-engineered due to IoT.

It seems to reason that tried-and-true approaches [9] may be revived to aid with the problem of managing large volumes of data given that the difficulties of handling enormous amounts of data has lasted for years. Data compression [20], "divide and conquer," "incremental learning," and "random sampling" [22, 23] are only a few of the approaches that can be found in the literature on data mining. Other methods include data compression [20], "divide and conquer," and "incremental learning." One possible solution [16] to the problem of large amounts of data that the Internet of Things presents is to have sensors capture just the valuable data rather than all of the data. The process of simplifying the data that is being collected has emerged as a central focus of study. Utilizing the principle component analysis (PCA) [24] or one of the many other approaches for dimension reduction is a straightforward way to cut down on the number of features present in the input data. Pattern reduction (PR) is a relatively recent

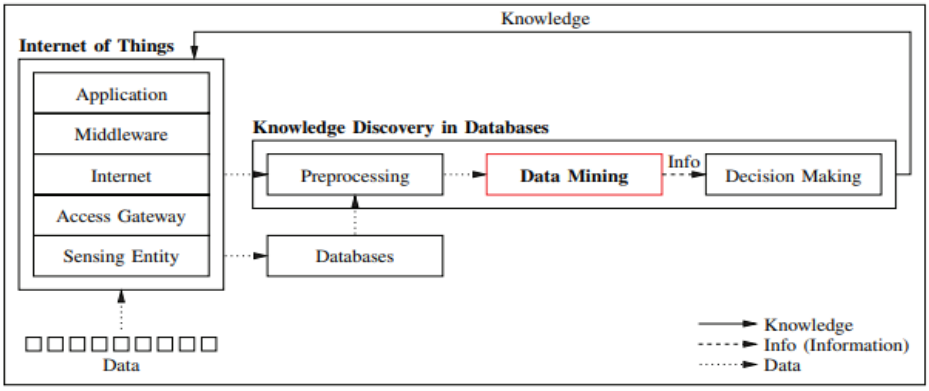
ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

concept that has been reported in [25, 26] and is based on a different viewpoint than traditional pattern recognition. In contrast to principle component analysis (PCA), this technique focuses on minimizing the number of patterns rather than the number of dimensions in an effort to cut down on the amount of time needed for the convergence process. As a result, it may be used to simplify the input data. In recent years, feature selection, distributed computing, and cloud computing have emerged as potential new paths for tackling the enormous data issue [27]. These techniques have been mostly superseded by these more modern approaches.

Figure 2 demonstrates that the Internet of Things collects data from a broad range of sources, some of which may potentially contribute data for the IoT. IoT data might potentially be transformed into actionable insights with the assistance of KDD, which would lead to new levels of comprehension. The output of the data processing phase is sent to the data mining phase, which then searches through it for meaningful patterns that may later be used in the decision-making phase. At any point in time throughout the mining process, the KDD procedure has the potential to have a major influence on the results. Because not all of the data's characteristics can be mined, feature selection is often used to choose the ones that are most relevant to each database entry. This is because not all of the attributes can be mined. Therefore, it is possible that data mining algorithms may have difficulty finding meaningful information if the selected features do not appropriately suit the attributes of the data (for instance, grouping patterns into appropriate groups). It is also crucial to note that KDD and data mining approaches may have less of an influence on the system performance and service quality of IoT applications in comparison to data fusion, big data, data transit, and distributed computing. This is something that should be taken into consideration.

Figure 2: The IoT architecture using KDD [63]



2.7. Data Mining for the internet of things (IoT)

In the next section, the connections between IoT data mining and KDD are addressed. A fundamental model of working is also provided to identify the relevant data mining (DM) technologies. A fundamental introduction to the IoT data mining technology is covered with a unified framework for data mining.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

The Basic idea to use DM for IoT: Data are considerably easier to create than data analysis. IoT will have a huge difficulty with the big expansion of data growth. So far different studies have been launched and the challenge of IoT large data processing has been resolved. There are insufficient efficient analytical instruments and all systems will therefore absolutely be combined with this vast quantity of data. From hardware perspective, if KDD is applied to IoT, the solution for huge data is cloud computing and certain other distributed technology. Most mining techniques are built and built-in order to work on one system from the point of view of the program. In the Big Data era it is impossible to directly use the most existing KDD system and typical data mining techniques to process that enormous volume of IoT data [64]. In general, KDD preprocessing procedures or data mining techniques need be updated to handle the massive amounts of data generated by IoT devices. On the other hand, established data mining algorithms can be applied to small-scale IoT systems that generate very little data [65].

The following three critical aspects in the KDD technology for the problem to be solved are used to design a high-performance data mining module for IoT: the purpose, data characteristics, and mining algorithm [66].

- **The objective (O):** The issue seeks to specify clearly if it must be addressed, with the problem's assumptions, restrictions, and measures established first. With such knowledge, the problem's goal may be stated very simply.
- **Data (D):** Data characteristics include size, distribution, and representation, which are the most essential aspects of DM. Various types of data require different processing methods. The data may originate from many applications and be connected to one another; however, if the data has distinct meanings, the data may be examined differently.
- **Mining methods (A):** If the above-mentioned purpose and data are well described, the data mining algorithm can be easily identified. A new mining algorithm must be devised, which cannot be easily justified based on the three concerns listed above. For example, if the volume of data exceeds a system's capability and there is no realistic way to regulate the data's complexity, a novel data mining methodology should be implemented immediately [66].

2.8. Data Mining for IoT Applications

IoT devices and Internet-connected sensors are rapidly growing; in this sector, we may discover various applications. The following lists some of the effective uses for data mining.

2.8.1. Smart city

The many IoT systems in a smart city are listed below, together with the relevant data mining functions, in order to make the system better and smarter[67].

- **Traffic Control:**

Traffic control IoT devices include the GPS system, smart phones, city-based vehicle sensors, which may supply data points such as journey duration, heavy and light frequency of cars, accident-prone locations and construction regions. The sensor data points give insights on the

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

reason for the congestion in the region. This scenario may be utilized to resolve the problem of traffic congestion using the classification method[68]. The target locations can be categorized according on the high, medium, low chance of traffic jams occurring in a certain location. The model may then be used to forecast the time of day after development of the classification model, at the height of traffic congestion. On this basis, the car might select the alternate path to the destination. Traffic is spread and the problem of congestion is avoided[69].

- **Residential electronic meters:**

Smart meters are currently quickly replaced by smart meters, and the intelligent meters are able to deliver real time energy usage statistics by email or digitally on smartphones. Throughout this time series data, a time series analysis data mining methodology may be employed, in which data is automatically obtained at different intervals for the entire day and may be used to anticipate energy consumption, and if an anomaly in energy consumption is discovered, it gives notifications in a variety of ways. This smart meter can also produce synthetic data from real-time data, which may be utilized for data forecasting [70].

- **Detection of pipelines leakage:**

Maintaining pipeline leaks is a laborious task for municipal companies. Sensors may be utilized particularly during the usage of old tubes to convey the sound of water through the tube utilizing data mining techniques. For leaks in the pipes, the external detection technique has been utilized. The application of this approach in the pipeline leakage can simplify the process of detection work for water leaks and also cut maintenance costs to half in comparison to standard methods of detection.

2.8.2. Automation at home

Many IoT devices in the home automation system can be used to generate useful information and patterns by extracted data generated by different sensors. The observed patterns can be used mainly for the prediction of future events and for better automated user engagement. Data mining models can be employed for the domestic automation system in classification and time series analysis [71]. Classification is employed by the classification of interactive gadgets which are closely connected and based on their use. Data mining methodology for time series analysis is utilized for the information generated by these devices with the matching time stamps, and linear regression may be employed in a certain time for this future event.

2.8.3. Health care

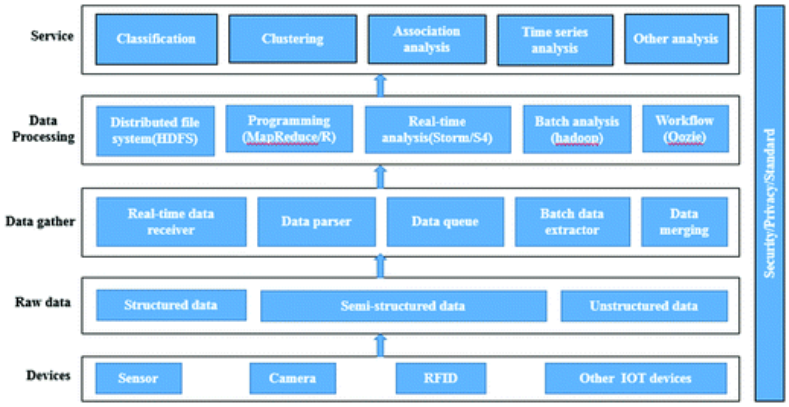
There is clearly evidence of the expansion of the healthcare industry because of the use and progress of IoT in this field. The IoT medical systems offer countless services for customers to check their health, such as medication adhesion systems, burnt calories, blood pressure, heart rate, blood glucose, weight measuring equipment and pulse measurements. An intelligent system should be designed for the integration of this heterogeneous data and for correct information about the patient [72]. The data can be mined by means of the patient's individual prescriptions and medical history, and we can draw key conclusions about the patient's current health and prospects of patient survival. Clustering the technology of data mining can be

ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

utilized with a view to improving patient care. Any strange patterns that are easy to detect fraud may also be investigated further.

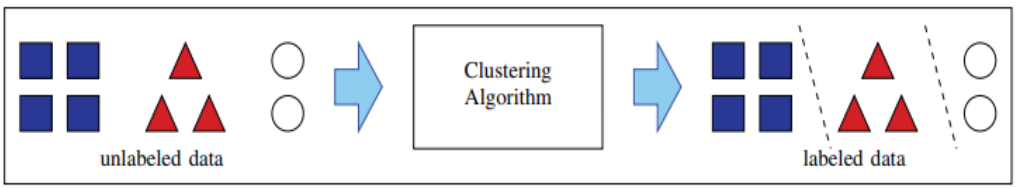
Figure 3: Data mining and IoT process



2.9. Data mining Clustering for IoT

Before attempting to analyze the outcomes of a clustering exercise, it is necessary to first implement a certain clustering approach to any appreciable degree. This is the fundamental tenet of clustering. Clustering measures such as sum of squared errors for data and peak signal-to-noise ratio (PSNR) for images are two examples of often used metrics [35]. The "K-means" technique is a well-known solution to difficult issues involving clustering information. To get started, the input patterns are segmented into their respective groups, and a set of centroids, denoted by the letter c, is generated to represent each of these groups. After that, the assignment operator will make use of the derived distances in order to place each pattern into the appropriate set. The task of this operator is to scan the input patterns and generate the applicant rules in order to ensure that each pattern is assigned to the appropriate group in the correct manner. After the assignment operator has finished assigning all of the patterns to their corresponding groups, the update operator may then alter the centers of those groups if necessary. k-means and its many variations [36] had an impact on the evolution of group algorithms throughout the course of time. An easy example that combines the two methods demonstrates how k-means may be used to enhance the results of other metaheuristic ways of clustering. [36] This is shown by the following: It is possible to overcome the challenge of soft clustering by using K-means and flouted logic, which ultimately results in improved clustering outcomes.

Figure 4: The process of clustering



The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

2.9.1. Clustering for IoT Infrastructure:

In recent years, numerous metaheuristic algorithms [73] utilize stochastic approaches to guess clustering outcomes, unlike classic clustering algorithms that apply deterministic local search methods for finding cluster outcomes (e.g., k-means). Due to the randomization characteristics of the solution searcher, metaheuristic algorithms are less likely in the early iterations to fall into the local optimum and hence are more likely to produce better results than local determinist search methods, notably in the case of big, complicated data settings. The accuracy is not the primary objective of the IoT clustering. The two other major objectives are to ensure that these new instances are clustered accordingly and that the problem requirements are met. A study of the clustering for the IoT focuses on determining the user comportment to offer the services required by the user. Another focus is distributed clustering [74], which is the fundamental necessity to extend the life of the Wireless Sensor Network (WSN). These observations lead to various innovative techniques of clustering.

2.9.2. Clusters for IoT Services:

Another study problem in the use of IoT clustering technologies means that IoT Services and applications make themselves decisions or offer better services such as a smart home system to detect elderly people's falling events. An interesting intelligent home was shown in [75], which focuses on the frequent patterns mining, identifying and following the user conduct of an intelligent house. The activity discovery method (ADM) is intended to cluse sequences based on a basic k-means algorithm after employing a discontinuous varied-order sequence miner (DVSM) to mine helpful patterns. However, the order and attribute distances utilized to quantify the sequences of frequent patterns in the system, rather than the Euclidean distance. The clustering algorithm takes the output of the frequent pattern mining technique as an input (DVSM in this case). The classification technique, which takes as an input a clustering result in identifying activities in such a smart environment based upon the hidden Markov model (HMM, a statistic model to describe relationships or probabilities between patterns).

2.9.3. Classification of internet of thinks (IoT)

Classification basic idea: Classification idea: as opposed to the clustering of a partitioning process which as a whole does not involve any prior knowledge, classification [76] [77] [78] requires some previous knowledge to direct the partitioning process in the development of a group of classifiers to reflect the possibility of pattern distribution. Classification is a supervised learning process, whereas clustering is an unsupervised learning process, as can be shown. The classification algorithm can be expressed mathematically as follows: Given a set of labeled data and a set of unlabeled data, the classifier (i.e., the hyperline or prediction function) is trained on the labeled data, and the unlabeled data is classified by the classifier.

2.9.4. Classification of IoT Infrastructures

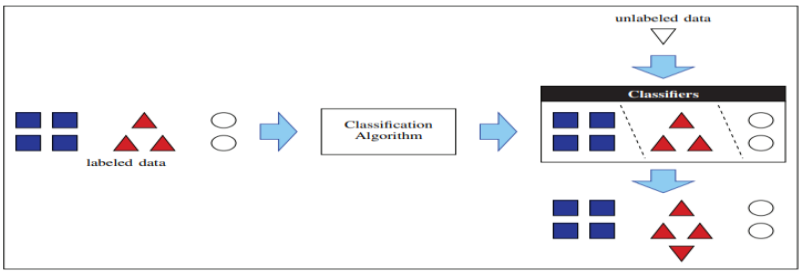
While IoT classification research is still in its infancy, it is nevertheless likely to improve the performance of IoT infrastructures and systems. One of the expected possibilities is the fact that numerous standards or applications created themselves in recent times as the UII (single

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

item identifier) schemes for the IoT are currently and in the future. The authors in [79] devised a binary tree classification technique for solving the problem with a view of avoiding or mitigating the issue of UII queries dramaticaly (even greater than the amount of DNS inquiries in the current internet environment). The tree-based categorization can quickly determine the device type by simply checking a number of the first bits of a unit. This means that the entire device header does not need to be checked.

Depending on the region, IoT categorization research can be classified into two categories: outdoor and inside. The gridlock problem, especially when you live in a big city, is one of the horrors constantly in our daily lives. There is more and more research into the traffic congestion problem by employing mobile devices like smartphones to exchange information to prevent traffic jamming [80] and accident sensing [81]. Since traffic information on the social networks and internet has become a trend [82], a driver guidance instrument was developed by merging the position information provided by the GPS, vehicle geographic data tracking and other information gathered from the internet in order to anticipate the future traffic scenario. This type of system then predicts and points the routing course of the driver based on real-time information, historical data, etc. The decision tree classification method is employed.

Figure 5: The process of data mining classification



2.10. Data mining frequent pattern for IoT

2.10.1. The basic idea of mining frequent pattern:

Mining for frequent patterns is notably distinct from several other mining approaches in two important areas. [41] [47]. To begin, the purpose of frequent pattern mining is to search through a huge dataset in order to uncover patterns that might be beneficial. The association rule is something that you may be acquainted with if you have experience with data mining. This approach is used to determine all of the connections that exist between a set of data items that have been gathered together. [48] [40] [49]. The idea that a simple example may be useful in explaining the notion of association rule is one that is often debated in the retail sector. Consider the common behavior of retailers, which is to often buy new things so that they may add them to the stock that they already have. The problem may be stated in mathematical terms as follows: Find the set of association rules that is equal to or greater than the predetermined threshold values of support and confidence, given a set of items (I = I1, I2,..., In), a set of transactions (T = (T1, T2,..., Tn) where TI I), and a preset threshold value of support and confidence. In other words, given this information, find the set of association rules. This suggests that the user has already decided that support and confidence are the two most

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

significant variables to consider when assessing the outcomes of mining. As an example, let's look at a transaction rule for purchasing bread and milk together, which is signified by the term "breadmilk." The support value will be 10%, and the confidence value will be 70%. According to this rule, only ten percent of consumers buy bread and milk at the same time, but if a client has previously purchased bread, there is a seventy percent chance that they will also purchase milk. Support, according to the mathematical definition, occurs when,

$$\text{support}(X \Rightarrow Y) = P(X \cup Y) = \frac{\text{TC}(X \cup Y)}{n},$$

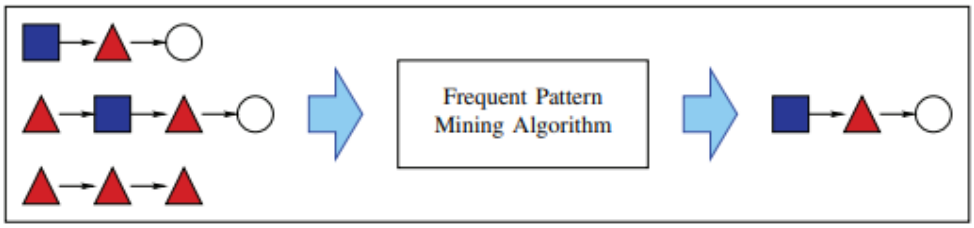
and the confidence is defined as

$$\text{confidence}(X \Rightarrow Y) = P(Y|X) = \frac{\text{TC}(X \cup Y)}{\text{TC}(X)},$$

Where TC(ψ) denotes the number of transactions in T that contains ψ, i.e., ψ = X ∪ Y or ψ = X.

Another difficulty that often comes up with frequent pattern mining is the "sequence" of transactions, referred to as sequential pattern [83] [77]. When it comes to association rules versus sequential patterns, the distinction is that an association rule is designed to discover interesting patterns from a transaction, while a sequential pattern is designed to discover interesting patterns from a succession of transactions. Consider the following statement: Sequential patterns may be described as follows. Given a collection of items I and a set of transactions T, the purpose of the sequential patterns problem is to identify all the sequences with a minimal support, support(s), where the minimal support of a sequence, support(s), is defined as the proportion of all the data sequences that contain the sequence s. The other three crucial factors we should follow while defining the scenario we want to study are the length of a time sequence, the time for events to be included in the time sequence, and the time interval for events to be included in the time sequence. Because the frequent pattern mining issue entails extracting hidden information from a database of transactions, various studies have sought to design more efficient algorithms for it, which resemble the apriori method [95]. The apriori method for the association rule issue aims to 1 scan the dataset for all the frequent item sets "fs" and 2 generate all the association rules from "fs" with a specific confidence value. The example in Figure 5 shows that, as the name implies, the goal of frequent pattern mining is to locate patterns that appear frequently in a database and fulfill preset support and confidence parameters.

Figure 6: The process of datamining frequent pattern [84]



The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

2.10.2. IoT's Frequent pattern of services

Analyzing purchasing behavior still attracts the attention of academics and firms employing RFID or even IoT approaches like most methods for the association rule. In [85] the researchers presented an excellent way of describing agents and consumer behavior, i.e. thinking, moving and doing, in a supermarket. To overcome the customer's failure to quickly identify the items that they want, customer specific regulations, category-based regulations and association rules are combined inside the same system of suggestions for purchases to help supermarket shoppers locate the items they want. Data are analyzed using rule based and case-based rationale in order to ensure that the system can propose what consumers require more precisely on the basis of 1 customer's personal benefit, 2 customer buys history and 3 others' conduct.

The smart environment, which needs mining technologies to make it smarter so as to give better services, is the second option based on RFID and sensor technology. A series of motions is utilized to allow classification algorithms to discriminate between various events unlike other classification studies that used everyday living activities (ADL) for describing the activities. The writers in [86] [87] [88] focused on temporal relationships. Multiple research [86]concentrated on characterizing and identifying the temporal activities relationship. The relationships are split in first, then, meetings, meetings, overlaps, overlaps, begins, starts, ends, ends during, includes and equals, which may then be utilized to explain relationships between temporal activities. A probability technique has been presented to estimate the chance of event X happening with another event Y as follows in order to forecast actions in an intelligent environment:

$$\begin{aligned} P(X|Y) = (&|\text{After}(Y,X)| + |\text{During}(Y,Z)| \\ &+ |\text{OverlappedBy}(Y,Z)| + |\text{MetBy}(Y,Z)| \\ &+ |\text{Starts}(Y,Z)| + |\text{StartedBy}(Y,Z)| \\ &+ |\text{Finishes}(Y,Z)| + |\text{FinishBy}(Y,Z)| \\ &+ |\text{Equals}(Y,Z)|)/|Y|. \end{aligned}$$

3. RELATED WORK

In recent years, a significant number of technological scholars have begun to use DM strategies that are founded on the Internet of Things. The following article is a research study that was conducted on the utilization of a variety of DM approaches in Internet of Things environments throughout the course of the last several years.

The authors of [54] present a unique Internet of Things-based wireless sensor network system that can analyze sensor data locally, without forwarding it to a central server or base station. They do this by using an association rule (BS). Because of the novel approach that was suggested, sensors will be able to carry out calculations locally, and only a condensed, high-quality statistical summary of the data will be required to be sent to the nodes that make up the cluster (CMs). This results in a reduction in the amount of needless communication and a large extension of the network's lifetime. When evaluating the suggested system, extensive

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

simulations are conducted, and the findings show that overhead communication may be considerably reduced by including the proposed scheme into already existing protocols, which in turn extends the network's life and stability.

[55] They presented a PADC (privacy and availability data clustering) strategy based on the algorithm and differential privacy as a means of enhancing initial center point selection and distance computation from additional locations to the center. Throughout the process of aggregation, PADC makes a concerted effort to lessen the significance of any data points that fall on the extreme end of the spectrum. The findings of the security review indicate that the procedure accomplishes the intended degree of granular privacy and prevents the leaking of sensitive data. In the meanwhile, a performance analysis demonstrates that the PADC scheme outperforms existing ways for providing intelligent Electrical Service in the IoT. This is due to the fact that it gives more clustering results while maintaining the same degree of privacy as current K-mean approaches.

Using the fundamentals of software engineering, the article [56] presents a framework that makes it possible to mine data from internet-connected devices. This structure was given the name (EdgeMiningSim). (EdgeMiningSim) This method guides domain experts in the direction of the disclosure of actionable data, such as descriptive or predictive models, which enables them to take the appropriate actions within the limited and ever-changing IoT environment. As a direct result of this, a case study was carried out with the help of a smart monitoring application in order to demonstrate a technique called (EdgeMiningSim) and emphasize the advantages of efficiently addressing all of the many elements that influence IoT data mining at the same time.

EdgeCloudSim, a framework for simulation, has been used in [57] to design both centralized and distributed K-Means clustering. The latter of these has been confirmed on a real system. By using an approximated and distributed clustering technique, one may achieve high levels of accuracy while still benefiting from reductions in the amount of energy used, the amount of time spent communicating, and the amount of computing required. Applications for the Internet of Things (IoT) that have requirements for low latency, low bandwidth/energy consumption, and minimal security and privacy issues may benefit tremendously from the ability of edge computing to perform data mining at the edge of the network. In addition to this, they used Edge CloudSim to quantitatively contrast the advantages of a Cloud-based strategy with those of an Edge-based strategy. While still retaining a high level of precision, approximate and dispersed clustering offers a number of useful improvements in terms of processing, communication, and the consumption of energy.

The authors of developed a method in [58] for classifying students into unique learning clusters on the basis of the particular studying practices that each student has. In the process of developing the architecture of the e-learning system, consideration was given to the unique capabilities of each student as a learner. The fundamental objective is to identify the best possible environments in which students may enhance their overall academic performance. By locating and putting into action a number of very important covert patterns, management may liberate the unrealized potential for change that the system has. In order to evaluate the

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

usefulness of educational data mining, techniques such as K-Medoids, Density-based Spatial Clustering of Noise Applications, Agglomerative Hierarchic Cluster Clustering, and Fast Search and Heater Diffusion Clustering "CFSFDP-HD" are used. The results indicated that the reliability of the findings improved when the "CFSFDP-HD" was used rather than the standard approaches. When it comes to analyzing enormous datasets for the goal of implementing educational regulations, data mining is a highly useful technique that may be used.

It was stated in [59] that one strategy for interpreting sensor data may comprise the following: (a dynamic data mining framework). A data-mining sensor model is built so that these oscillations may be used to their full potential. There are many different applications for sensor networks, each of which represents a different kind of physical system. The gathering and examination of sensor data changes from the past may provide us with information on the qualities of a physical system. Capitalizing on the "associations" that exist between the properties of the different physical systems allows for the discovery of connections that exist between the various configurations of sensor networks. In this controlled laboratory environment, examples of physical parameters that were being measured were transfer distance, transmission latency, sensor data, data changes, and so on. Experiments have been carried out on the particular experimental platform, and the results show that the model is able to mine dynamic data and find stable patterns. The platform was selected for this purpose. The outcomes of the tests were used to provide a benchmark for dynamic data mining to the model, and new strategies for assessing large amounts of data in industry were devised as a consequence. In [60], two fundamental models for modeling sensor data are presented: the LSTM prediction model and the Support Vector Machine (SVM) model based on IoT data. Both models are flexible to a broad variety of data volumes. Extensive simulations validate Hadoop's exceptional character in terms of prediction accuracy and training effectiveness across a diverse array of operational scenarios. This nature is proven by the fact that Hadoop can adapt to a broad variety of operational settings.

The provision of a method for keeping tabs on a patient's vital signs is the objective of [61]. (a case of cardiac arrhythmia). When testing and running the system that monitors heart rhythm and conducts electrocardiography, an Arduino board and an AD8232 sensor module were used as the primary components. The k-Nearest Neighbor (KNN) technique, which is one of the most used data mining methodologies, is utilized in order to categorize and validate the kind of cardiac arrhythmia that a patient is experiencing. As a consequence of this, it has been shown that the approach that was proposed is capable of classifying and confirming the kind of cardiac arrhythmia with a high level of precision. In [62] An EAA-SMO approach for sequential minimum optimization (SMO) is the goal of this study. The EAA is an enhanced version of the original Apriori algorithm that takes into account knowledge of the relevant context. Mining consumer data allows healthcare firms to provide superior service to their patients. When it comes to evaluating other less visible or concealed data models, health data mining offers a plethora of fresh choices to consider. In addition, the created rule is classified as a prediction model for anomaly detection that is based on SMO regression. This classification was applied to the rule after it was developed. The outcomes of the experiments indicate that the suggested

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

method achieves a higher level of accuracy (by 2%), and it operates 25% more quickly than the baseline (when compared to semantic ontology).

The authors of [63] provide a detailed and well-organized examination of the numerous data mining approaches that are used to nurture an intelligent setting. These methods may be employed in large-scale IoT projects as well as small-scale IoT initiatives. The process of data mining in today's data-driven world is supported by advanced classification and clustering technologies, which function as the process's backbone. Support vector machines, hidden Markov models, Bayesian networks, logistic regression, deep neural networks, k-means, and DBSCAN are a few examples of the types of algorithms that may be used. Discovering sequential real-time events or patterns requires the use of a number of different algorithms, some of which include FP-growth, Episode Discovery, Varied-order Sequential Miner, Deep Belief Network, and variations. Each of these algorithms contributes significantly to the process. For the future infrastructure of the Internet of Things, expertise in extensive research and development as well as machine learning will be necessary.

They referred to missing data for medical IoT applications in their paper [64] and developed a hybrid neural network that was combined with a genetic algorithm. We employ a model that is constructed from a deep learning neural network (Jordan network), and the weights are modified using a genetic algorithm. This allows us to anticipate missing data. According to the results, the suggested strategy has the potential to boost the overall performance of an IoT application by up to 5% by assuming missing data that has a high classification score based on area under the curve (AUC). The authors of [65] investigated the properties of wireless communication as well as its underlying technologies by situating their research within the context of the Internet of Things. A multi-tree data mining strategy was suggested, which resulted in an improvement to the wireless communication mechanism that is based on the Internet of Things. In addition, a system is constructed, and tests are carried out in order to assess the data mining approach. The findings demonstrate that the data mining method was successful. Improving the wireless connection mechanism via the use of the Internet of Things is the best option available when compared to other data mining tactics that are on par with one another.

In order to model complex data abstraction, they examined machine learning and artificial neural networks (ANN), and in order to enhance IoT data, they looked at six prominent information mining approaches [66]. C4.5 and C5.0 are the most accurate and trustworthy of the group based on the results of the testing done on a real-world smarter IoT dataset. ANNs are capable of producing accurate results while preserving the computational elegance.

The energy-saving grid [67] makes extensive use of clustering techniques as an important component. In order to choose a cluster head in the most effective manner, it was necessary to concentrate on a process that would transport CH around among the nodes that had comparatively greater energy levels (CH). The algorithm takes into account the starting energy, the residual energy, and the optimal value of CHs in order to choose the following set of CHs for the network when it is used for IoT applications such as environmental monitoring, intelligent city systems, and others. According to the results of the simulation research, the

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

improved version is superior than the protocol for adaptive clustering in terms of throughput, lifespan, and residual power by a respective 60%, 66%, and 64%. Academics published their Apriori methodology in [68], and due to the fact that Apache Spark is equipped with in-memory processing capabilities, its implementation of the technique is superior to that of the MapReduce framework.

Because this technique capitalizes on qualities that are shared by both K-Means and K-Harmonic Mean, it was recommended in [69] that a new hybrid methodology be created that uses MapReduce to combine K-Means with K-Harmonic Mean in order to generate more exact conclusions. In contrast to the K-Means algorithm, which generates local optimum results that are below par, the K-Harmonic Mean algorithm improves upon these metrics while simultaneously producing a greater number of them. As a direct result of this, the proposed hybrid algorithm performs superiorly than its competitors in the context of local best-case scenarios.

In reference number 70, the author develops a decentralized ensemble approach by making use of SmoothSVM, a fast support vector machine (SVM) methodology. When compared to the use of a single SVM, they were effective in developing and evaluating a large ensemble of SVMs in simultaneously with a little amount of additional work required. It is possible that training an ensemble of SVMs will take less time than training a single SVM, with test accuracy that is equivalent to or even superior. One other advantage is that it is not difficult to scale up to much more extensive systems. Data mining approaches, namely a clustering algorithm, were applied, as the authors in [71] mentioned, in order to increase the performance of the network.

The sensitivity and specificity of an IDS (Intrusion Detection System) in were evaluated with the use of three different classifiers: the Naive Bayes, the J48, and the Random Forest. The KDD NSL dataset is used here in order to carry out the tests. The research results demonstrate that when accuracy and detection rate are compared, Random Forest comes out on top, beating out both J48 and Naive Bayes. [73] conducted an analysis of studies that analyzed the mining of opinions for mobile app stores. A unique convolutional neural network (CNN) on document representation was trained with the use of the first dataset of its type to contain both genuine and spam assessments of applications. This dataset was utilized to train the CNN. When compared to a baseline classification model based on a Support Vector Machine (SVM), which obtained only 70% accuracy with different feature combinations, the technique based on neural networks achieved 82.5% accuracy. Alternately, the author in [74] gave a performance of data mining algorithms that safeguard mobile devices from hackers. These techniques included the Naive Bayes Multiclass classifier, J48, SVM, Random Forest, and Decision Tree. As a direct consequence of this, the Naive Bayes technique is superior to the other approach in terms of both the speed and accuracy of detection.

In order to enhance the functionality of the first AprioriAll sequence mining algorithm described in [75], research into data mining technology with enormous IoT processing and a new Binary AprioriAll sequence mining approach were carried out, and both were put into use. Because of the adoption of the method, there is a noticeable improvement in the effectiveness of the database. Because of the reduction in the amount of work that has to be done, the

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

efficiency of the system has improved. Researchers in [76] used a technique called maximum frequent patterns (MFP) of opcode sequences in order to differentiate malicious IoT applications from benign ones. Because of this, they are able to identify malware related to internet of things devices with a success rate of 99%. In the [77] study, the use of a data mining strategy was presented as a method for obtaining monitoring of the driving conditions of vehicles in terms of wheel speed and GPS data. The Random Forest Classification is one of the most used ways to data mining. It has been shown to have an accuracy rate of 80.9% after being researched.

According to the ideas presented in [78], the smart gateway concept of fog computing might be used to perform remote monitoring of patients' health. The solution that has been suggested makes use of modern technologies and approaches, such as embedded data mining, distributed storage, and real-time alerting, amongst others. An event-based system is used for the transfer of data, and this allows the processing of patient data on a fog layer to take place in real time. By determining the patient's temporal health index, which is analyzed through temporal mining, unfavorable occurrences may be better understood. For a period of one month, 67 patients living in smart homes that were equipped with Internet of Things technology collected health data that demonstrated the effectiveness of the system. When compared to alternative classification strategies, it has been shown that the Bayesian belief network classification model, which is the one that is suggested, has greater accuracy and reaction time. Incorporating data about the patient's current state of health into the decision-making process is another way to significantly improve the provided remedy.

Table 1: Comparison of Datamining Techniques with IoT

Ref.	Year	Mining techniques	Work on	Result
[89]	2021	Apriori Algorithm	Apriori algorithm used with MapReduce framework and Apache Spark implementations	Apache Spark implementations of Apriori show better performance due to in memory processing capabilities
[90]	2018	Clustering (k-mean) with K-Harmonic Mean	suggests a novel hybrid technique that combines K-Means and K-Harmonic Mean by utilizing MapReduce in order to get more accurate results since this technique makes advantage of aspects of both methods	Because K-Means has a poor local optimum, the K-Harmonic Mean solves this problem while also improving local optimal search results. As a result, the suggested hybrid algorithm performs better in local optimal situations.
[91]	2019	SmoothSVM SVM	SmoothSVM, a rapid support vector machine (SVM) technique, was employed to create a distributed ensemble technique. When compared to a single SVM, they were able to generate and assess a large ensemble of SVMs in parallel with low overhead.	The ensemble of SVMs may be trained in less time than a single SVM while keeping the same test accuracy or even improving it in some circumstances. It also has the added benefit of scaling to far bigger systems with ease.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[92]	2019	Clustering	They suggested wireless sensor to work on	As a result, shown that the algorithm performed a better for network performance enhancement
[93]	2018	Naive Bayes Random Forest J48	They examined the detection rate and accuracy of IDS (Intrusion Detection System) using Naive Bayes, J48, and Random Forest classifiers. The KDD NSL dataset is utilized in the experiments	The results indicate that Random Forest outperformed Naive Bayes and J48 in terms of accuracy and detection rate.
[94]	2019	Classification (SVM) CNN	The mobile app store opinion mining studies offered to be evaluated. Used a new CNN (convolutional neural network) to learn documents representation by characterizing the app store review dataset that contains for the first time in literature genuine and spam reviews.	The neural network-based technique obtained 82.5% accuracy, while a baseline Support Vector Machine (SVM) classification model attained just 70% accuracy with diverse feature combinations.
[95]	2018	Naïve Bayes Multiclass classifier J48, SVM Random Forest, Decision Tree	The Performance of data mining algorithms to secure mobile devices from malicious attacks	The outcome is that Naive Bayes is far superior in time and also in detection than the other methods.
[96]	2013	AprioriAll algorithm	The original AprioriAll sequence mining algorithm was enhanced by a study of data mining technology with huge IOT processing and a new Binary AprioriAll sequence mining method.	As the result the algorithm improves the efficiency of the database. This reduces the burden on the system, improving the efficiency.
[97]	2020	KNN, SVM, AdaBoost, decision tree, Random Forest, MLP	To distinguish malicious from benign IoT apps, researchers used detected maximum frequent patterns (MFP) of opcode sequences	As a result, they achieve an accuracy rate of 99% in the detection of unseen IoT malware.
[98]	2019	Random forest.	Using a data mining algorithm to obtain monitoring of vehicle driving conditions in terms of wheel speed and GPS data	One of the most often used data mining techniques is Random Forest Classification. The precision gained by the investigation of the random forest classification is 80.9 per cent.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[99]	2020	Cluster WSN system	proposing a new IoT-based WSN system to reduce sensor data by association rule without transferring it to any Cluster head (CH) or base station (BS). The novel suggested system allows sensors to do local calculations and just a minimal, superior statistical summary of the data is sent among cluster members (CMs).	the findings show that integrating the suggested scheme in current protocols substantially decreases the overhead communication which eventually extends the life and stability of the network
[100]	2018	Clustering "PADC" K-means	propose a privacy and availability data clustering "PADC" strategy based on the k-means algorithm and differential privacy that improves the initial center point selection as well as the distance calculation approach from other points to the center point.	In comparison to current differential privacy k-means algorithms, performance evaluation demonstrates that the proposed "PADC" scheme enhances the availability of clustering results at the same privacy level, implying that the proposed "PADC" scheme outperforms others for intelligent electrical service in IoT.
[101]	2021	Clustering EdgeMiningSim	they proposed the (EdgeMiningSim,) a software engineering principles-based framework for allowing IoT data mining. (EdgeMiningSim) The field experts are guided by this methodology to reveal practical information, namely descriptive or predictive models for appropriate action in the restricted and dynamic IoT context	as a result, shown that a case study is installed with a smart monitoring application that aims to demonstrate an (EdgeMiningSim) method and demonstrate the advantages of dealing successfully with all the multi-faceted elements that impact IoT data mining simultaneously.
[102]	2019	K-means clustering EdgeCloudSim	Centralized and distributed K-Means' clustering have been built in the EdgeCloudSim simulation framework and validated on a real system. Selecting an approximated and distributed clustering approach can bring benefits in terms of compute, communication, and energy usage, while retaining high levels of accuracy	In practice, approximated and dispersed clustering provides benefits in terms of computing, communication, and energy usage, while retaining high levels of accuracy.
[103]	2018	K-Medoids "CFSFDP-HD"	they Presented a student partition grouping methodology based on their learning behavior in different groups or clusters. In order to recognize and respond to the content in accordance with the	The substitution of conventional approaches with "CFSFDP-HD" was seen to generate more robust results. The technology of data mining is similarly efficient in

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

			learning capacity of students, a personalized e-learning system architecture is provided.	analyzing massive data to enforce education systems.
[104]	2020	Mining association Sensor data mining	provide a dynamic data mining framework for sensor data processing. A data extraction sensor model is created that may be employed in the dynamic change process. Various network contexts of sensors are considered separate physical systems in approach. The physical system and its parameters are trained in the collection and extraction of historical changes in sensor data.	Experiments on the selected experimental platform have been performed and the results show that the model can mine the dynamic data and identify stable patterns. By analyzing experimental findings, the model had a reference value for dynamic data mining, and new approaches for industrial big data analysis had been developed.
[105]	2020	k-Nearest Neighbor (KNN) (AD8232) sensor	offer a method for monitoring a patient's health (a case of cardiac arrhythmia). The system to monitor heart rhythm and perform electrocardiography was tested and ran using an Arduino circuit board and an (AD8232) sensor module. As a result, one of the most widely used data mining methods, the (k-Nearest Neighbor) technique, is utilized to categorize and validate the kind of cardiac arrhythmia.	As a result, shown that the proposed algorithm has a high accuracy classify and validate the type of cardiac arrhythmia
[106]	2019	LSTM SVM	they provided two fundamental models for sensor data modeling: the LSTM prediction model and the Support Vector Machine (SVM) model based on IoT data, both of which are suited for diverse data quantities.	Extensive simulations are run to confirm Hadoop's exceptional nature in terms of a high prediction accuracy and better training efficiency under a variety of operating conditions.
[107]	2020	Apriori algorithm (EAA) (SMO)	For sequential minimum optimization (SMO), an enhanced Apriori algorithm (EAA) based on context ontology knowledge (EAA-SMO) technique is proposed. Healthcare businesses can benefit from data mining.	When compared to semantic ontology, the suggested method improves accuracy by 2% and reduces execution time by 25%, according to the results of the experiments.

ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

[108]	2020	SVM, HMM, Bayesian network, logistic regression, DNN, k-mean, and DBSCAN	provided a systematic and in-depth examination of numerous data mining techniques used in big and small-scale IoT applications to create an intelligent environment. The most often used data mining methods are advanced classification and clustering techniques (SVM, HMM, Bayesian network, logistic regression, DNN, k-mean, and DBSCAN	they found that Future IoT infrastructure will need extensive research and development, as well as machine learning capabilities
[109]	2019	Deep neural network (DNN)	To refers missing data for medical IoT applications, they presented a hybrid neural network with genetic algorithm. To anticipate missing data, a deep learning neural network (Jordan network) is utilized as a model, with the weights of the neural network optimized using a genetic method.	he results show that the suggested technique may impute missing data with a high classification value based on Area Under the Curve (AUC) and improve the final performance of an IoT application by up to 5%.
[110]	2020	Multi-tree data mining technique (MLDM)	they looked at the features and essential technologies of wireless communication in the context of the Internet of Things. Proposed a multi-tree data mining technique (MLDM), which improved the wireless communication mechanism based on the Internet of Things in real time.	The results indicate that the data mining method performs well. Among the similar data mining techniques, it is preferable to improve the wireless communication mechanism based on IoT, as it is more practicable and useful.
[111]	2019	Artificial neural network (ANN)	They looked at six well-known information mining methods for better IoT data, as well as machine learning and artificial neural networks (ANN) for modeling complex data abstraction	The experimental results on a real smarter IoT dataset show that C4.5 and C5.0 have better accuracy and dependability than the other algorithms. ANNs offer very exact findings while remaining elegant in their computation.
[112]	2019	Cluster head (CH) Clustering algorithm	The focus was on an efficient choice system that rotates CH among the nodes with greater energy levels in contrast to others while selecting a cluster head (CH). The algorithm takes into account beginning energy, residual energy, and the optimal value of CHs to choose the following group of CHs for the network which is appropriate for IoT applications such as environmental monitoring,	as a result, the improved version works more effectively than the protocol for adaptive clustering by improving throughput by 60 per cent of 66 per cent of lifespan and 64 per cent of residual power.

ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

			intelligent city systems and other systems	
[113]	2018	state-of-the-art technology Bayesian belief network	They suggested remote monitoring of patients' health at smart homes utilizing the smart gateway idea of fog computing. The suggested approach utilizes state-of-the-art technology and services, such as embedded data mining, distributed storage and on-line notification. To demonstrate the validity of the system, health data were systematically produced for 30 days from 67 patients in an intelligent home environment based on IoT.	Results show the high accuracy and reaction time in determining the state of an event when compared to other classification methods in the suggested (Bayesian belief network) classification model.

4. DISCUSSION

In spite of the fact that it may seem to be impossible, the Internet of Things (IoT) will, in the not-too-distant future, make a great number of things that were once considered to be impossible feasible. One example of this is the ability to link every device on the planet to the internet. Because of this, there will be a considerable adjustment made to the way that we go about our daily lives. The mining of data is without a doubt going to be an important part of the process of making such technology intelligent enough to provide more pleasurable services and settings, and that part is what I'm referring to. In the next section of the essay, we will discuss the work of a few authors who have accomplished great things via the use of data mining and the Internet of Things. To provide one example, the authors of "73" have penned a number of articles on the topic of opinion mining for app stores. These articles are available on the internet for your perusal. When compared to the baseline Support Vector Machine (SVM) classification model, which only achieved 70% accuracy with various feature combinations, we discovered that using a novel CNN (convolutional neural network) to learn the representation of documents helped us achieve 82.5% accuracy. This was achieved in comparison to the SVM classification model, which only achieved 70% accuracy with various feature combinations. In spite of the fact that the SVM classification model was used, this was still the outcome. In the annals of academic research, one thing that has never been done before

ISSN 1533-9211

DOI 10.17605/OSF.IO/3QRVX

is to compile a dataset of app store reviews that include both genuine ratings and fake evaluations, sometimes known as spam reviews. In addition to being mentioned in other publications, the authors are cited in [77]. The authors are also mentioned in other publications. The Random Forest Classification method is one of the most widely used ways to data mining. This method has been used in order to acquire monitoring of the driving conditions of autos in terms of wheel speed and GPS data. Through analysis of the random forest classification approach, we were able to achieve improvements in accuracy that amounted to 80.9%.

5. CONCLUSION

Data Mining (DM) technologies for the Internet of Things are discussed in this study (IoT). so, IoT has become a hot study topic, as it promises to improve quality of life and safety in Smart Cities, increase resource supply and waste management, and optimize traffic, among other things. We evaluated studies on applying data mining technologies to the IoT in this study, which includes classification, clustering, and frequent patterns mining technologies, from the infrastructure and service perspectives. There is also an examination and discussion of the scale of each data mining technology as well as the overall integrated system. To help the readership of the paper completely comprehend the changes brought about by the Internet of Things, Finally, this article discussed and concluded the data mining applications with internet of things.

References

[1] T. M. G. Sami, Z. S. Ageed, Z. N. Rashid, and Y. S. Jghef, "Distributed, Cloud, and Fog Computing Motivations on Improving Security and Privacy of Internet of Things," Mathematical Statistician and Engineering Applications, vol. 71, pp. 7630-7660, 2022.

[2] N. R. Omar, R. H. Saeed, J. A. Ahmed, S. B. Muhammad, Z. S. Ageed, and Z. N. Rashid, "Enhancing OS Memory Management Performance: A."

[3] S. H. Haji, S. Zeebaree, R. H. Saeed, S. Y. Ameen, H. M. Shukur, N. Omar, et al., "Comparison of software defined networking with traditional networking," Asian Journal of Research in Computer Science, pp. 1-18, 2021.

[4] H. M. Yasin, S. Zeebaree, M. Sadeeq, S. Y. Ameen, I. M. Ibrahim, R. R. Zebari, et al., "IoT and ICT based smart water management, monitoring and controlling system: A review," Asian Journal of Research in Computer Science, vol. 8, pp. 42-56, 2021.

[5] R. J. Hassan, S. Zeebaree, S. Y. Ameen, S. F. Kak, M. Sadeeq, Z. S. Ageed, et al., "State of art survey for iot effects on smart city technology: challenges, opportunities, and solutions," Asian Journal of Research in Computer Science, vol. 22, pp. 32-48, 2021.

[6] Z. S. AGEED, Z. N. RASHID, R. T. HASAN, Y. S. JGHEF, T. M. G. SAMI, and M. A. OMER, "WEB TECHNOLOGY DEVELOPMENT AND DATA-INTENSIVE INFLUENCES ON SEMANTIC INFORMATION RETRIEVING BASED ON PARALLEL AND CLOUD COMPUTING."

[7] S. Zeebaree, S. Ameen, and M. Sadeeq, "Social media networks security threats, risks and recommendation: A case study in the kurdistan region," International Journal of Innovation, Creativity and Change, vol. 13, pp. 349-365, 2020.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[8] H. R. Abdulqadir, S. R. Zeebaree, H. M. Shukur, M. M. Sadeeq, B. W. Salim, A. A. Salih, et al., "A study of moving from cloud computing to fog computing," Qubahan Academic Journal, vol. 1, pp. 60-70, 2021.

[9] L. M. Abdulrahman, S. Zeebaree, S. F. Kak, M. Sadeeq, A. 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.

[10] F. E. F. Samann, S. R. Zeebaree, and S. Askar, "IoT provisioning QoS based on cloud and fog computing," Journal of Applied Science and Technology Trends, vol. 2, pp. 29-40, 2021.

[11] L. M. Haji, O. M. Ahmad, S. Zeebaree, H. I. Dino, R. R. Zebari, and H. M. Shukur, "Impact of cloud computing and internet of things on the future internet," Technology Reports of Kansai University, vol. 62, pp. 2179-2190, 2020.

[12] K. Jacksi, S. R. Zeebaree, and N. Dimililer, "LOD Explorer: Presenting the Web of Data," International Journal of Advanced Computer Science and Applications, vol. 9, 2018.

[13] A. AL-Zebari, S. Zeebaree, K. Jacksi, and A. Selamat, "ELMS–DPU ontology visualization with Protégé VOWL and Web VOWL," Journal of Advanced Research in Dynamic and Control Systems, vol. 11, pp. 478-85, 2019.

[14] A. Zeebaree, A. Adel, K. Jacksi, and A. Selamat, "Designing an ontology of E-learning system for duhok polytechnic university using protégé OWL tool," J Adv Res Dyn Control Syst Vol, vol. 11, pp. 24-37, 2019.

[15] H. S. Yahia, S. Zeebaree, M. Sadeeq, N. Salim, S. F. Kak, A. Adel, et al., "Comprehensive survey for cloud computing based nature-inspired algorithms optimization scheduling," Asian Journal of Research in Computer Science, vol. 8, pp. 1-16, 2021.

[16] K. Jacksi, N. Dimililer, and S. R. Zeebaree, "A survey of exploratory search systems based on LOD resources," 2015.

[17] H. M. Yasin, S. R. Zeebaree, and I. M. Zebari, "Arduino based automatic irrigation system: Monitoring and SMS controlling," in 2019 4th Scientific International Conference Najaf (SICN), 2019, pp. 109-114.

[18] D. A. Zebari, H. Haron, S. R. Zeebaree, and D. Q. Zeebaree, "Multi-level of DNA encryption technique based on DNA arithmetic and biological operations," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 312-317.

[19] S. Zebari and N. O. Yaseen, "Effects of parallel processing implementation on balanced load-division depending on distributed memory systems," J. Univ. Anbar Pure Sci, vol. 5, pp. 50-56, 2011.

[20] K. B. Obaid, S. Zeebaree, and O. M. Ahmed, "Deep learning models based on image classification: a review," International Journal of Science and Business, vol. 4, pp. 75-81, 2020.

[21] H. I. Dino, S. R. Zeebaree, D. A. Hasan, M. B. Abdulrazzaq, L. M. Haji, and H. M. Shukur, "COVID-19 diagnosis systems based on deep convolutional neural networks techniques: A review," in 2020 International Conference on Advanced Science and Engineering (ICOASE), 2020, pp. 184-189.

[22] Z. J. Hamad and S. R. Zeebaree, "Recourses utilization in a distributed system: A review," International Journal of Science and Business, vol. 5, pp. 42-53, 2021.

[23] S. Zebari, "A new approach for process monitoring," Polytechnic Journal, Technical Education-Erbil, 2011.

[24] S. M. Saleem, S. R. Zeebaree, and M. B. Abdulrazzaq, "Real-life dynamic facial expression recognition: a review," in Journal of Physics: Conference Series, 2021, p. 012010.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[25] B. W. Salim and S. R. Zeebaree, "Design & Analyses of a Novel Real Time Kurdish Sign Language for Kurdish Text and Sound Translation System," in 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), 2020, pp. 348-352.

[26] R. Ibrahim, S. R. Zeebaree, and K. Jacksi, "Semantic Similarity for Document Clustering using TFIDF and K-mean," Master's Thesis, 2020.

[27] S. A. Mostafa, A. Mustapha, A. A. Ramli, R. Darman, S. R. Zeebaree, M. A. Mohammed, et al., "Applying Trajectory Tracking and Positioning Techniques for Real-time Autonomous Flight Performance Assessment of UAV Systems," Journal of Southwest Jiaotong University, vol. 54, 2019.

[28] G. M. O. Zebari, K. Faraj, and S. Zeebaree, "Hand Writing Code-PHP or Wire Shark Ready Application Over Tier Architecture with Windows Servers Operating Systems or Linux Server Operating Systems," 2016.

[29] R. K. Ibrahim, S. R. Zeebaree, K. Jacksi, M. A. Sadeeq, H. M. Shukur, and A. Alkhayyat, "Clustering Document based Semantic Similarity System using TFIDF and K-Mean," in 2021 International Conference on Advanced Computer Applications (ACA), 2021, pp. 28-33.

[30] A. S. Abdulraheem, S. R. Zeebaree, and A. M. Abdulazeez, "Design and implementation of electronic human resource management system for duhok polytechnic university," Technology Reports of Kansai University, vol. 62, pp. 1407-1420, 2020.

[31] G. M. Zebari, S. Zeebaree, M. M. Sadeeq, and R. Zebari, "Predicting Football Outcomes by Using Poisson Model: Applied to Spanish Primera División," Journal of Applied Science and Technology Trends, vol. 2, pp. 105-112, 2021.

[32] S. H. Ahmed and S. Zeebaree, "A survey on Security and Privacy Challenges in Smarthome based IoT," NEW ARCH-INTERNATIONAL JOURNAL OF CONTEMPORARY ARCHITECTURE, vol. 8, pp. 489-510, 2021.

[33] H. H. A. Razaq, A. S. Gaser, M. A. Mohammed, E. T. Yassen, S. A. Mostafad, S. R. Zeebaree, et al., "Designing and implementing an arabic programming language for teaching pupils," Journal of Southwest Jiaotong University, vol. 54, 2019.

[34] H. Abdullah and S. R. Zeebaree, "Android Mobile Applications Vulnerabilities and Prevention Methods: A Review," 2021 2nd Information Technology To Enhance e-learning and Other Application (IT-ELA), pp. 148-153, 2021.

[35] Z. S. Ageed, S. R. Zeebaree, M. A. Sadeeq, R. K. Ibrahim, H. M. Shukur, and A. Alkhayyat, "Comprehensive Study of Moving from Grid and Cloud Computing Through Fog and Edge Computing towards Dew Computing," in 2021 4th International Iraqi Conference on Engineering Technology and Their Applications (IICETA), 2021, pp. 68-74.

[36] S. Zeebaree, N. Cavus, and D. Zebari, "Digital Logic Circuits Reduction: A Binary Decision Diagram Based Approach," LAP LAMBERT Academic Publishing, 2016.

[37] Y. S. Jghef, M. J. M. Jasim, H. M. Ghanimi, A. D. Algarni, N. F. Soliman, W. El-Shafai, et al., "Bio-Inspired Dynamic Trust and Congestion-Aware Zone-Based Secured Internet of Drone Things (SIoDT)," Drones, vol. 6, p. 337, 2022.

[38] S. Chavhan, S. R. Zeebaree, A. Alkhayyat, and S. Kumar, "Design of Space Efficient Electric Vehicle Charging Infrastructure Integration Impact on Power Grid Network," Mathematics, vol. 10, p. 3450, 2022.

[39] N. T. Muhammed, S. R. Zeebaree, and Z. N. Rashid, "Distributed Cloud Computing and Mobile Cloud Computing: A Review," QALAAI ZANIST JOURNAL, vol. 7, pp. 1183-1201, 2022.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[40] L. M. Abdulrahman, S. R. Zeebaree, and N. Omar, "State of Art Survey for Designing and Implementing Regional Tourism Web based Systems," Academic Journal of Nawroz University, vol. 11, pp. 100-112, 2022.

[41] A. S. Aljuboury, S. R. Zeebaree, F. Abedi, Z. S. Hashim, R. Q. Malik, I. K. Ibraheem, et al., "A New Nonlinear Controller Design for a TCP/AQM Network Based on Modified Active Disturbance Rejection Control," Complexity, vol. 2022, 2022.

[42] R. K. Ibrahim, S. R. Zeebaree, K. Jacksi, S. H. Ahmed, S. M. Mohammed, R. R. Zebari, et al., "Clustering Document based on Semantic Similarity Using Graph Base Spectral Algorithm," in 2022 5th International Conference on Engineering Technology and its Applications (IICETA), 2022, pp. 254-259.

[43] H. B. Abdalla, A. M. Ahmed, S. R. Zeebaree, A. Alkhayyat, and B. Ihnaini, "Rider weed deep residual network-based incremental model for text classification using multidimensional features and MapReduce," PeerJ Computer Science, vol. 8, p. e937, 2022.

[44] K. P. M. Kumar, J. Mahilraj, D. Swathi, R. Rajavarman, S. R. Zeebaree, R. R. Zebari, et al., "Privacy Preserving Blockchain with Optimal Deep Learning Model for Smart Cities," CMC-COMPUTERS MATERIALS & CONTINUA, vol. 73, pp. 5299-5314, 2022.

[45] A. D. Kshemkalyani and M. Singhal, Distributed computing: principles, algorithms, and systems: Cambridge University Press, 2011.

[46] A. Sunyaev, Internet computing: Principles of Distributed systems and emerging internet-based technologies: Springer Nature, 2020.

[47] G. Dudek and M. Jenkin, Computational principles of mobile robotics: Cambridge university press, 2010.

[48] V. D. Majety, N. Sharmili, C. R. Pattanaik, E. L. Lydia, S. R. Zeebaree, S. N. Mahmood, et al., "Ensemble of Handcrafted and Deep Learning Model for Histopathological Image Classification," CMC-COMPUTERS MATERIALS & CONTINUA, vol. 73, pp. 4393-4406, 2022.

[49] A. M. Abed, Z. N. Rashid, F. Abedi, S. R. Zeebaree, M. A. Sahib, A. J. a. Mohamad Jawad, et al., "Trajectory tracking of differential drive mobile robots using fractional-order proportional-integral-derivative controller design tuned by an enhanced fruit fly optimization," Measurement and Control, p. 00202940221092134, 2022.

[50] Z. N. Rashid, S. R. Zeebaree, M. A. Sadeeq, R. R. Zebari, H. M. Shukur, and A. Alkhayyat, "Cloud-based Parallel Computing System Via Single-Client Multi-Hash Single-Server Multi-Thread," in 2021 International Conference on Advance of Sustainable Engineering and its Application (ICASEA), 2021, pp. 59-64.

[51] Z. N. Rashid, S. R. Zeebaree, and A. Sengur, "Novel remote parallel processing code-breaker system via cloud computing," ed: TRKU, 2020.

[52] F. Abedi, S. R. Zeebaree, Z. S. Ageed, H. M. Ghanimi, A. Alkhayyat, M. A. Sadeeq, et al., "Severity Based Light-Weight Encryption Model for Secure Medical Information System."

[53] B. W. SALIM and S. R. ZEEBAREE, "ISOLATED AND CONTINUOUS HAND GESTURE RECOGNITION BASED ON DEEP LEARNING: A REVIEW."

[54] R. Buyya, J. Broberg, and A. M. Goscinski, Cloud computing: Principles and paradigms: John Wiley & Sons, 2010.

[55] H. Attiya and J. Welch, Distributed computing: fundamentals, simulations, and advanced topics vol. 19: John Wiley & Sons, 2004.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[56] A. P. Jardosh, K. N. Ramachandran, K. C. Almeroth, and E. M. Belding-Royer, "Understanding congestion in IEEE 802.11 b wireless networks," in Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, 2005, pp. 25-25.

[57] A. Nicolaou, N. Temene, C. Sergiou, C. Georgiou, and V. Vassiliou, "Utilizing mobile nodes for congestion control in wireless sensor networks," in 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2019, pp. 1-7.

[58] S. R. M. Zeebaree, "New Hashing Algorithm and an Authentication Technique to Improve IoT Security."

[59] D. M. Abdullah and S. R. Zeebaree, "Comprehensive survey of IoT based arduino applications in healthcare monitoring."

[60] S. Vasudevan, B. DeCleene, N. Immerman, J. Kurose, and D. Towsley, "Leader election algorithms for wireless ad hoc networks," in Proceedings DARPA Information Survivability Conference and Exposition, 2003, pp. 261-272.

[61] N. A. Ali and M. Abu-Elkheir, "Data management for the internet of things: Green directions," in 2012 IEEE Globecom Workshops, 2012, pp. 386-390.

[62] S. B. B. Priyadarshini, A. BhusanBagjadab, and B. K. Mishra, "The role of IoT and big data in modern technological arena: A comprehensive study," in Internet of Things and Big Data Analytics for Smart Generation, ed: Springer, 2019, pp. 13-25.

[63] J. F. De Paz, J. Bajo, V. F. López, and J. M. Corchado, "Biomedic Organizations: An intelligent dynamic architecture for KDD," Information Sciences, vol. 224, pp. 49-61, 2013.

[64] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, and L. T. Yang, "Data mining for internet of things: A survey," IEEE Communications Surveys & Tutorials, vol. 16, pp. 77-97, 2013.

[65] S. Choudhary and N. Kesswani, "Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 datasets using deep learning in IoT," Procedia Computer Science, vol. 167, pp. 1561-1573, 2020.

[66] C.-W. Tsai, C.-F. Lai, and A. V. Vasilakos, "Future internet of things: open issues and challenges," Wireless Networks, vol. 20, pp. 2201-2217, 2014.

[67] B. P. L. Lau, S. H. Marakkalage, Y. Zhou, N. U. Hassan, C. Yuen, M. Zhang, et al., "A survey of data fusion in smart city applications," Information Fusion, vol. 52, pp. 357-374, 2019.

[68] M. M. Rashid, J. Kamruzzaman, M. M. Hassan, S. S. Shafin, and M. Z. A. Bhuiyan, "A survey on behavioral pattern mining from sensor data in Internet of Things," IEEE Access, vol. 8, pp. 33318-33341, 2020.

[69] M. Sattarian, J. Rezazadeh, R. Farahbakhsh, and A. Bagheri, "Indoor navigation systems based on data mining techniques in internet of things: a survey," Wireless Networks, vol. 25, pp. 1385-1402, 2019.

[70] J.-S. Chou, Y.-C. Hsu, and L.-T. Lin, "Smart meter monitoring and data mining techniques for predicting refrigeration system performance," Expert Systems with Applications, vol. 41, pp. 2144-2156, 2014.

[71] V. H. Bhide and S. Wagh, "i-learning IoT: An intelligent self learning system for home automation using IoT," in 2015 International Conference on Communications and Signal Processing (ICCSP), 2015, pp. 1763-1767.

[72] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," in 2018 15th learning and technology conference (L&T), 2018, pp. 29-34.

[73] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," ACM computing surveys (CSUR), vol. 35, pp. 268-308, 2003.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[74] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach," in IEEE INFOCOM 2004, 2004.

[75] P. Rashidi, D. J. Cook, L. B. Holder, and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment," IEEE transactions on knowledge and data engineering, vol. 23, pp. 527-539, 2010.

[76] M. Kantardzic, Data mining: concepts, models, methods, and algorithms: John Wiley & Sons, 2011.

[77] B. Liu, Web data mining: exploring hyperlinks, contents, and usage data: Springer Science & Business Media, 2007.

[78] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, et al., "Top 10 algorithms in data mining," Knowledge and information systems, vol. 14, pp. 1-37, 2008.

[79] Y. H. Lee, H. J. Kim, B.-h. Roh, S. W. Yoo, and Y. Oh, "Tree-based classification algorithm for heterogeneous unique item ID schemes," in International Conference on Embedded and Ubiquitous Computing, 2005, pp. 1078-1087.

[80] E. J. Horvitz, J. Apacible, R. Sarin, and L. Liao, "Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service," arXiv preprint arXiv:1207.1352, 2012.

[81] C. Thompson, J. White, B. Dougherty, A. Albright, and D. C. Schmidt, "Using smartphones to detect car accidents and provide situational awareness to emergency responders," in International Conference on Mobile Wireless Middleware, Operating Systems, and Applications, 2010, pp. 29-42.

[82] K. Perera and D. Dias, "An intelligent driver guidance tool using location based services," in Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, 2011, pp. 246-251.

[83] Q. Zhao and S. S. Bhowmick, "Sequential pattern mining: A survey," ITechnical Report CAIS Nayang Technological University Singapore, vol. 1, p. 135, 2003.

[84] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach," Data mining and knowledge discovery, vol. 8, pp. 53-87, 2004.

[85] C. Schwenke, V. Vasyutynskyy, and K. Kabitzsch, "Simulation and analysis of buying behavior in supermarkets," in 2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETFA 2010), 2010, pp. 1-4.

[86] V. Jakkula, D. J. Cook, and A. S. Crandall, "Temporal pattern discovery for anomaly detection in a smart home," 2007.

[87] G. Papamatthaiakis, G. C. Polyzos, and G. Xylomenos, "Monitoring and modeling simple everyday activities of the elderly at home," in 2010 7th IEEE Consumer Communications and Networking Conference, 2010, pp. 1-5.

[88] S. Lühr, G. West, and S. Venkatesh, "Recognition of emergent human behaviour in a smart home: A data mining approach," Pervasive and Mobile Computing, vol. 3, pp. 95-116, 2007.

[89] S. Raj, D. Ramesh, and K. K. Sethi, "A Spark-based Apriori algorithm with reduced shuffle overhead," The Journal of Supercomputing, vol. 77, pp. 133-151, 2021.

[90] U. Bagde and P. Tripathi, "An analytic survey on mapreduce based k-means and its hybrid clustering algorithms," in 2018 second international conference on computing methodologies and communication (iccmc), 2018, pp. 32-36.

[91] J. Hajewski and S. Oliveira, "Distributed SmSVM ensemble learning," in INNS big data and deep learning conference, 2019, pp. 7-16.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[92] N. T. Hanh, H. T. T. Binh, N. X. Hoai, and M. S. Palaniswami, "An efficient genetic algorithm for maximizing area coverage in wireless sensor networks," Information Sciences, vol. 488, pp. 58-75, 2019.

[93] N. Ashraf, W. Ahmad, and R. Ashraf, "A comparative study of data mining algorithms for high detection rate in intrusion detection system," Annals of Emerging Technologies in Computing (AETiC), Print ISSN, pp. 2516-0281, 2018.

[94] N. Genc, "Detection of spam review on mobile app stores, evaluation of helpfulness of user reviews and extraction of quality aspects using machine learning techniques," École de technologie supérieure, 2019.

[95] K. Uma and E. S. Blessie, "Survey on Android Malware Detection and Protection using Data Mining Algorithms," in 2018 2nd 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 2nd International Conference on, 2018, pp. 209-212.

[96] J. Ji, W. Wang, W. Yan, and Z. Tong, "Contrast and study on binary AprioriAll sequence mining algorithm of IOT massive information processing," in Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, 2013, pp. 467-470.

[97] H. Darabian, A. Dehghantanha, S. Hashemi, S. Homayoun, and K. K. R. Choo, "An opcode-based technique for polymorphic Internet of Things malware detection," Concurrency and Computation: Practice and Experience, vol. 32, p. e5173, 2020.

[98] J. H. Joon, D. A. Renata, S. H. Yoon, C. Youngjune, and Y. Z. Jembre, "Data Mining for Speed Bump Detection from Car Wheels and GPS Sensors using Random Forest," in 2019 International Conference on Information and Communication Technology Convergence (ICTC), 2019, pp. 740-743.

[99] A. M. Khedr, W. Osamy, A. Salim, and S. Abbas, "A Novel Association Rule-Based Data Mining Approach for Internet of Things Based Wireless Sensor Networks," IEEE Access, vol. 8, pp. 151574-151588, 2020.

[100] J. Xiong, J. Ren, L. Chen, Z. Yao, M. Lin, D. Wu, et al., "Enhancing privacy and availability for data clustering in intelligent electrical service of IoT," IEEE Internet of Things Journal, vol. 6, pp. 1530-1540, 2018.

[101] C. Savaglio and G. Fortino, "A Simulation-driven Methodology for IoT Data Mining Based on Edge Computing," ACM Transactions on Internet Technology (TOIT), vol. 21, pp. 1-22, 2021.

[102] C. Savaglio, P. Gerace, G. Di Fatta, and G. Fortino, "Data mining at the IoT edge," in 2019 28th International Conference on Computer Communication and Networks (ICCCN), 2019, pp. 1-6.

[103] S. Kausar, X. Huahu, I. Hussain, Z. Wenhao, and M. Zahid, "Integration of data mining clustering approach in the personalized E-learning system," IEEE Access, vol. 6, pp. 72724-72734, 2018.

[104] Y. Yin, L. Long, and X. Deng, "Dynamic data mining of sensor data," IEEE Access, vol. 8, pp. 41637-41648, 2020.

[105] E. Moghadas, J. Rezazadeh, and R. Farahbakhsh, "An IoT patient monitoring based on fog computing and data mining: Cardiac arrhythmia usecase," Internet of Things, vol. 11, p. 100251, 2020.

[106] Y. Zheng and G. Chen, "Energy Analysis and Application of Data Mining Algorithms for Internet of Things Based on Hadoop Cloud Platform," IEEE Access, vol. 7, pp. 183195-183206, 2019.

[107] Y. Yan, C. Huang, Q. Wang, and B. Hu, "Data mining of customer choice behavior in internet of things within relationship network," International Journal of Information Management, vol. 50, pp. 566-574, 2020.

The Seybold Report
ISSN 1533-9211
DOI 10.17605/OSF.IO/3QRVX

[108] P. Sunhare, R. R. Chowdhary, and M. K. Chattopadhyay, "Internet of Things and Data Mining: An Applications Oriented Survey," Journal of King Saud University-Computer and Information Sciences, 2020.

[109] N. Al-Milli and W. Almobaideen, "Hybrid neural network to impute missing data for iot applications," in 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019, pp. 121-125.

[110] L. Li, "Real time auxiliary data mining method for wireless communication mechanism optimization based on Internet of things system," Computer Communications, vol. 160, pp. 333-341, 2020.

[111] M. Mahendra, C. Kishore, and C. Prathima, "Data mining efficiency and scalability for smarter internet of things," in Computational Intelligence and Big Data Analytics, ed: Springer, 2019, pp. 119-125.

[112] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand, and A. H. Gandomi, "Residual energy-based cluster-head selection in WSNs for IoT application," IEEE Internet of Things Journal, vol. 6, pp. 5132-5139, 2019.

[113] P. Verma and S. K. Sood, "Fog assisted-IoT enabled patient health monitoring in smart homes," IEEE Internet of Things Journal, vol. 5, pp. 1789-1796, 2018.