REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technologies

AN EXHAUSTIVE AND METHODICAL ANALYSIS OF LITERATURE PERTAINING TO TECHNIQUES FOR MANAGING BIG DATA WITHIN THE CONTEXT OF THE INTERNET OF THINGS

Dhiaa Mahdi Jaseem JASEEM

Master's Thesis

Supervisor

Assoc. Prof. Dr. Sefer KURNAZ

İstanbul, 2024

AN EXHAUSTIVE AND METHODICAL ANALYSIS OF LITERATURE PERTAINING TO TECHNIQUES FOR MANAGING BIG DATA WITHIN THE CONTEXT OF THE INTERNET OF THINGS

Dhiaa Mahdi Jaseem JASEEM

Information Technologies

Master’s Thesis

ALTINBAŞ UNIVERSITY

2024

The thesis titled AN EXHAUSTIVE AND METHODICAL ANALYSIS OF LITERATURE PERTAINING TO TECHNIQUES FOR MANAGING BIG DATA WITHIN THE CONTEXT OF THE INTERNET OF THINGS prepared by DHIAA MAHDI JASEEM and submitted on 6/ 06/2024 has been **accepted unanimously** for the degree of Master of Science in Information Technologies.

Assoc. Prof. Dr. Sefer URNAZ

Supervisor

Thesis Defense Jury Members:

Assoc. Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altinbas University	___________________
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altinbas University	___________________
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical Engineering,, Istanbul Arel University	___________________

I hereby declare that this thesis meets all format and submission requirements of a Master`s Thesis.

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Dhiaa Mahdi Jaseem JASEEM

Signature

DEDICATION

I would like to dedicate this work to my dearest Mom, Dad and my beloved wife. I feel so honored and blessed to have you all by my side and want to express my gratitude for your care and support over the years, thanks for instilling me with a strong passion for learning and for doing everything possible to put me on the path to greatness, I will never forget the important values you have passed down to me particularly perseverance and honesty word cannot describe how important you are to me.

ACKNOWLEDGEMENTS

I might want to thank my administrators: Dr. Sefer KURNAZ Please let me express my profound feeling of appreciation and gratefulness of you for the information, direction and unrestricted help you have given me. I want you to enjoy all that life has to offer and further achievement and accomplishments throughout your life.

ABSTRACT

AN EXHAUSTIVE AND METHODICAL ANALYSIS OF LITERATURE PERTAINING TO TECHNIQUES FOR MANAGING BIG DATA WITHIN THE CONTEXT OF THE INTERNET OF THINGS

JASEEM, Dhiaa Mahdi Jaseem

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Assoc. Prof. Dr. Sefer KURNAZ

Date: June / 2024

Pages: 58

Big Data and the Internet of Things (IoT) have brought about previously unheard-of opportunities and challenges for data management. The body of research on big data management techniques in the context of the Internet of Things is thoroughly and methodically reviewed in this study. The study covers a broad range of research, from fundamental ideas to innovative techniques. With the goal of enhancing user convenience and lifestyle, the Internet of Things (IoT) has emerged as a transformative force. It incorporates a number of underlying technologies, including networking, sensor technologies, and man-to-machine and machine-to-machine communication. The Internet of Things' success depends on these technologies' ability to manage data effectively. This article addresses a range of data-related topics, including data sources, collection methods, processing strategies, and transmission devices, in order to investigate the challenges and issues associated with managing information in the setting of the Internet of Things. The article describes and discusses the difficulties posed by having to handle massive volumes of heterogeneous data across multiple systems. These difficulties are discussed in terms of the logical and practical components of data management and communication networks. In addition, the paper looks at IoT data models and techniques for data control, cleaning, and indexing that take into consideration the unique characteristics of IoT data. In-depth consideration of the benefits and limitations of data management in the Internet of Things is provided in the article's concluding sections. Real-world applications are presented to illustrate the practical implications of these considerations, offering insights into the ways IoT data management can positively impact various domains.

Keywords: IoT, Big Data, BDA, Big Data and IoT, Big Data Challenges.

ÖZET

NESNELERİN İNTERNETİ BAĞLAMINDA BÜYÜK VERİYİ YÖNETME TEKNİKLERİNE İLİŞKİN LİTERATÜRÜN KAPSAMLI VE METODOTİK ANALİZİ

JASEEM, Dhiaa Mahdi Jaseem

Yüksek Lisans., Bilgi Teknolojileri, Altınbaş Üniversitesi

Danışman: Doç. Dr. Sefer KURNAZ

Tarih: June / 2024

Sayfa: 58

Büyük Veri ve Nesnelerin İnterneti (IoT), veri yönetimi için daha önce duyulmamış fırsatları ve zorlukları beraberinde getirdi. Bu çalışmada Nesnelerin İnterneti bağlamında büyük veri yönetimi teknikleri üzerine yapılan araştırmalar kapsamlı ve metodik olarak gözden geçirilmektedir. Çalışma, temel fikirlerden yenilikçi tekniklere kadar geniş bir araştırma yelpazesini kapsıyor. Kullanıcı rahatlığını ve yaşam tarzını geliştirme hedefiyle Nesnelerin İnterneti (IoT), dönüştürücü bir güç olarak ortaya çıktı. Ağ oluşturma, sensör teknolojileri ve insandan makineye ve makineden makineye iletişim dahil olmak üzere bir dizi temel teknolojiyi içerir. Nesnelerin İnterneti'nin başarısı, bu teknolojilerin verileri etkili bir şekilde yönetme becerisine bağlıdır. Bu makale, Nesnelerin İnterneti ortamında bilgi yönetimiyle ilgili zorlukları ve sorunları araştırmak amacıyla veri kaynakları, toplama yöntemleri, işleme stratejileri ve iletim cihazları dahil olmak üzere veriyle ilgili bir dizi konuyu ele almaktadır. Makalede, çok sayıda sistemdeki büyük hacimli heterojen verileri işlemek zorunda kalmanın yarattığı zorluklar açıklanıyor ve tartışılıyor. Bu zorluklar veri yönetimi ve iletişim ağlarının mantıksal ve pratik bileşenleri açısından tartışılmaktadır. Ek olarak makale, IoT verilerinin benzersiz özelliklerini dikkate alan IoT veri modellerine ve veri kontrolü, temizleme ve indeksleme tekniklerine de bakmaktadır. Makalenin sonuç bölümlerinde Nesnelerin İnterneti'nde veri yönetiminin yararları ve sınırlamaları derinlemesine ele alınmaktadır. Bu hususların pratik sonuçlarını göstermek için gerçek dünya uygulamaları sunulur ve IoT veri yönetiminin çeşitli alanları olumlu yönde etkileyebileceği yollara ilişkin bilgiler sunulur.

Anahtar Kelimeler: IoT, Büyük Veri, BDA, Büyük Veri ve IoT, Büyük Veri Zorlukları.

TABLE OF CONTENTS

Pages

LIST OF TABLES

LIST OF FIGURES

Pages

ABBREVIATIONS

IoT	:	Internet Of Things
BDA	:	Big Data Analytics
ICT	:	Information And Communication Technology
CAGR	:	Compound Annual Growth Rate
AI	:	Artificial Intelligence
DB	:	Data Base
IP	:	Internet Protocol
SQL	:	Structured Query Language
RDBMS	:	Relational Database Management Systems
CAGR	:	Compound Annual Growth Rate
DS	:	Discovery Service
RTRS	:	Real-Time Response Systems
CEP	:	Complex Event Processing
STMS	:	Smart Traffic Management System
IIOT:	:	Industrial Internet of Things
MMBD:	:	Multimedia Big Data
HPC	:	HPC
RFID	:	RFID
RTOS	:	Real-Time Operating System.

1. INTRODUCTION

The widespread use of interconnected devices has fundamentally transformed the development, collecting, and processing of data, resulting in the emergence of significant difficulties related to big data. This article examines the intersection between big data and the Internet of Things, which is a cross-disciplinary problem. The objective is to offer readers a thorough comprehension of the technologies responsible for managing the vast quantities of data produced by Internet of Things devices. Big data, a very influential and pervasive invention in the contemporary world, has rapidly spread across the globe. The multiplicity of contexts in which data is found, such as social media, the internet of Things (IoT) devices, smart towns, and sectors of industry, facilitates the proliferation of data sources. The data in big data is complex and vast in quantity since it is varied and heterogeneous. Conventional database systems are incapable of managing large data because of its vast volume, high speed, and resistance to conform to pre-existing database structures. In order to derive value from this data, other processing methods are necessary. Big data, as defined by Merv Adrian, refers to data that beyond the capability of existing software and technology in terms of its volume, intricacy, and processing speed. Regarding the McKinsey Global Institute, these datasets are more capable of handling, storing, processing, and collecting information than standard database processing techniques can handle. Big data is generated from data volumes that exceed the capabilities of current technological systems to effectively handle, store, and process.

In order to maximise energy efficiency, electric power engineers must allocate substantial efforts towards enhancing energy consumption projections and optimisation. Smart cities are the result of several technology solutions that aim to enhance operational efficiency, service provision, and integrated municipal management. Cities that are smart aim to function in a coordinated and intelligent manner through the use of technology in the areas of housing, infrastructure, medical care, schooling, transportation, electricity, and environmental protection. Furthermore, this enhances the public's access to information and elevates the quality of services, while also promoting greater awareness, involvement, and efficiency. In order to address the difficulties caused by rising urbanisation, the use of information and communication technologies (ICTs) is crucial in promoting sustainable development policies and ensuring their global implementation. The notion of a "smart city" has gained popularity, leading the administrations of some major cities to collect extensive data in order to derive valuable insights. These concepts can assist individuals in enhancing their standard of living and uniting in support of initiatives that foster enduring sustainability. To enhance the living circumstances and level of life for residents, it is necessary to reduce expenses and optimise energy use. Cost-cutting measures targeted at increasing efficiency have the potential to significantly benefit industries like education, transportation, emergency medical care, and emergency services. The extensive utilisation of massive data storage systems is a direct result of smart grid technology. Public institutions, such as those in charge of water and electricity distribution, now regularly monitor the energy usage of their clients. In order to maintain energy efficiency in the face of obstacles like population growth and the rise in energy consumption brought on by new technologies, a smart grid—which is closely related to the Internet of Things—is required.

Smart grid technologies have been suggested as a means of modernising electricity systems. In order to do this, it is necessary for the various elements involved in electricity generation, transmission, and distribution to function in conjunction with communication networks, computers, and sensors. This connectivity facilitates the gathering of data, the management of supply and demand, and the prediction of energy use, hence simplifying these processes. Smart metres facilitate the collecting of real-time demand data and offer precise predictions of future demand levels. Customers receive comprehensive data regarding their energy usage, while utility suppliers obtain even more information through the interconnected metres. During moments of high demand, demand response mechanisms such as load dynamics and knowledge accumulation can be utilised. Network equipment sensors not only enhance network efficiency and reliability, but also save operational expenses by swiftly detecting and resolving errors. The widespread use of smart grids, however, presents additional obstacles that need the utilisation of big data technology to evaluate vast quantities of data.

Many firms extensively depend on big data analytics, which have a substantial impact on decision-making processes. The objective of this article is to offer companies and consumers precise assessments of energy production and consumption across a network. This is achieved by providing the findings of a study conducted at the district level, which utilised data from electricity metres to analyse energy usage. The dataset includes the weather data, acknowledging its significant influence on consumption rates. The K-Means Algorithm is utilised to cluster weather-related variables, including temperature, dew point, UV index, and others, in order to produce unique weather situations. Gaining insight into the relationship between weather and consumption patterns is essential for effectively resolving issues with energy networks, and this clustering technique facilitates the acquisition of such knowledge. The aim of this project is to measure energy consumption rates and evaluate the effect of weather on consumption in order to collect vital information required for fixing problems with the energy network.

A thorough examination of big data management methodologies inside the IoT framework reveals a complex ecosystem characterised by several challenges and corresponding solutions. The copious volumes of data produced by IoT devices require efficient approaches for gathering, storing, processing, evaluating, and deriving insights. Ongoing research is being performed to address the specific demands and limitations of managing Big Data in Internet of Things (IoT) environments. This research encompasses several techniques, including data fusion, machine learning, stream processing, distributed computing, and predictive analytics. Moreover, specialists emphasise the need of data privacy, scalability, security, and interoperability in guaranteeing the effective implementation and utilisation of these strategies. A comprehensive literature evaluation can assist academics in advancing theory and practice by shedding light on present practices, emerging trends, and potential future directions for managing Big Data in the Internet of Things (IoT) ecosystem.

Although data analytics has been around for a while, big data analytics (BDA) has gained prominence as a method to analyse the enormous volumes of unorganised data produced by the Internet of Things (IoT). Buckland et al. (2016) predict that the number of Internet of Things (IoT) connections will reach 27 billion by 2025, with a Compound Annual Growth Rate (CAGR) of 16 percent. In addition, they expect that the amount of data generated by IoT devices would exceed 2 zettabytes by the given year. According to Scales (2016), the estimated revenue from the Internet of Things is expected to reach USD 3 trillion by 2025. This is a substantial increase from USD 750 billion in 2015. The overall revenue is predicted to exceed $1.3 trillion, with end-user products (devices, connections, and apps) contributing to this amount. The Internet of Things will account for the remaining USD 1.3 trillion, encompassing application development, system integration, hosting, and data monetization.

According to Geng (2017), as we move closer to the fourth industrial revolution, the Internet of Things (IoT) will have a big impact on people's lives as well as how businesses operate. The components of Internet of Things devices, as listed by Holdowsky et al. (2015), include sensors, networks, standards, augmented intelligence, and augmented behaviour.

The integration of IoT devices is vital for organisations because to the revolutionary benefits it provides, such as enhanced forecasting and predictive analytics. Multiple data sources, including as GPS data, mobile sensors, and GIS-based visual analytics, together with real-time workflow analysis and event stream processing, are necessary for this (Behmann and Wu, 2015). The primary result of integrating next-generation analytics driven by IoT devices is the creation of a flexible environment that enables quick decision-making based on perceptive interpretations of events and behaviours.

Buyya and Dastjerdi (2016) categorise all advanced analytics derived from IoT devices into two types of data: big stream and big data. "Big data" refers to permanent records that are maintained in a centralised cloud, whereas "big stream" or "tiny data" refers to the more temporary recordings that are often acquired by Internet of Things (IoT) smart devices.

Figure 1.1 demonstrates the notable disparities between the current state of affairs and the initial phases of IoT implementation. The exorbitant cost and significantly restricted memory, storage, and performance of computer networks were the main factors contributing to these discrepancies. In contrast, contemporary Internet of Things (IoT) devices provide increased storage and memory capabilities at more cost-effective rates.

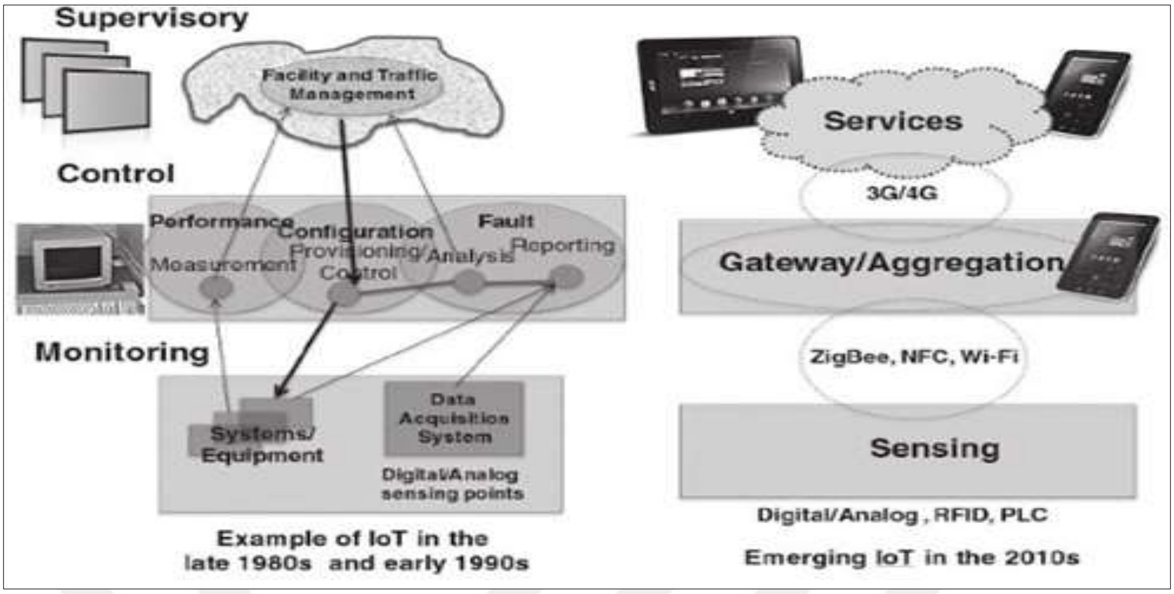

Figure 1.1: Disparities Between IoT Functional Models of the Past and Present.

These enhancements provide aggregated statistics and allow businesses to monetize IoT data on a larger scale. Consequently, we are seeing the dawning of a new era marked by the introduction of services like big data and analytical data on Sundays. Grabis and Kirikova (2011) identified cooperation, event-based interaction, and dynamic adaptability as three essential components of the Internet of Things ecosystem. In distributed systems, it enables the flexible interconnection of components and collaborates to efficiently complete the required tasks. In addition, he consistently handles a wide range of events originating from physical environments, actively configuring tools and platforms for collaboration to facilitate the implementation of ideas. The focus of this work is to examine how PDA (Personal Digital Assistant) differs from typical analytics in IoT (Internet of Things) scenarios. The aim is to create a classification system for IoT and IoT environments in order to understand the transformations and disruptions that the use of IoT and IoT devices has brought about in different industries and sectors. This article aims to analyse the possible applications of the Internet of Things (IoT). Explore the emerging possibilities offered by the Internet of Things, namely in the areas of instantaneous data streaming and event analysis. Examine the main concerns regarding the safety, confidentiality, and interoperability of the Internet of Things. Provide an overview of the proposed frameworks, solutions, and structures.

Examine the integration of mobile and cloud assistive technologies inside the newly implemented Internet of Things initiatives. Gain expertise in diverse data mining and business intelligence approaches. This study provides a comprehensive analysis of Big Data generated by Internet of Things (IoT) devices, with a particular focus on their real-time operations. The content is partitioned into five primary segments: The text includes a descriptive analysis of articles, research on problems related to database system relationships and the emergence of real-time analysis, current data and events in Internet of Things systems, discussion of real-time analysis using techniques and methods for extracting big data, and useful insights from the Internet of Things. This text provides a description of the study methods and findings of a systematic review, as well as the measures taken in analysing large data from the Internet of Things. After analysing the various components, the article examines the security, privacy, and interoperability challenges in Internet of Things systems and discusses how these challenges led to the development of various frameworks, architectures, and solutions. The subsequent section focuses on the significance of mobile and cloud assistive technologies, particularly highlighting their ongoing advancements within the Internet of Things. In conclusion, the article encompasses a debate, recommendations for further investigation, and a final summary.

1.1 DATA IN THE INTERNET OF THINGS

The Internet of Things (IoT) refers to a system of networked computing devices, sensors, and physical things that may exchange data in real time across the internet. These gadgets encompass a wide range of products, including smart household appliances, fitness trackers, and machinery and equipment utilized in factories.

Since its introduction, the Internet of Things has experienced significant growth. The Internet of Things enables the collection and analysis of data from real items, offering several benefits to individuals and businesses. The Internet of Things enhances people's lives by equipping ordinary things with "smart" functionalities that enable them to function autonomously with minimal human interaction. For instance, smart thermostats provide accurate regulation of temperature, while smart outlets aid in monitoring energy use. IoT sensors oversee industrial equipment and notify enterprises when repairs are needed, while intricate logistical systems enhance fuel efficiency.

Information on clients, regular duties, and internal organizational procedures may easily occupy terabytes of storage capacity. The data is obtained from a diverse range of origins, such as intelligent sensors, mobile devices, vehicles, and intelligent machinery. This paper extensively examines the techniques used to gather, transmit, store, and analyses vast quantities of data.

a. For the Internet of Things to operate and be used, data is essential. The Internet of Things' main idea is that by connecting numerous objects to the internet, data can be independently collected, sent, and shared by them. The data is derived from the embedded technology of these devices, encompassing components such as sensors and actuators. The data ecosystem serves the aim of collecting and managing data in relation to the Internet of Things. One of its key elements is data acquisition, where sensors embedded in IoT devices are specifically intended to receive data from their immediate environment. These sensors can pick up on a number of different things, such as motion, humidity, temperature, pressure, and intensity of light. These sensors enable people and organizations to actively monitor, assess, and react to changes in real time by providing insightful data about the physical world.

b. Data transfer: After being gathered, data from the Internet of Things (IoT) is sent to centralized servers, cloud platforms, or other networked devices locally through networks or globally via the internet. Such communication can be conducted by various wireless and wired networks, such as power lines, Wi-Fi, Bluetooth, and ZigBee. In order to achieve fast and consistent interaction, data processing and decoding systems inside the Internet of Things must possess the capability to effectively transfer data.

c. Data processing is the third phase, when IoT data is analyses to uncover valuable insights and suggestions. These actions occur at many stages in the data transmission process, encompassing hybrid configurations, cloud platforms, the edge, or even on the device itself. By processing and analyzing data directly on the device, edge computing reduces latency, preserves bandwidth, and improves privacy and security.

d. Data storage: Preserving IoT data is crucial for ensuring compliance, doing analysis, and maintaining records for future reference. The optimal storage option is determined by the amount, speed, and diversity of data, and might range from a local device or gateway to a distributed database, data warehouse, or cloud service. Concerns linked to data storage encompass issues such as the ability to handle increasing amounts of data, the dependability of the storage system, the regulations for how long data is retained, and adherence to laws governing the protection of data.

e. Data analysis: Harnessing data from the Internet of Things yields useful insights that improve decision-making, optimisation, and innovation. Advanced analytical techniques, including anomaly detection, machine learning, predictive analytics, and artificial intelligence, can be employed to reveal trends, patterns, correlations, and outliers in data streams produced by the Internet of Things (IoT). This data has the potential to assist firms in identifying novel revenue streams, enhancing their operational efficiency, and enhancing customer satisfaction.

f. Safeguarding the confidentiality and integrity of data: Ensuring the security of IoT data is crucial to protect it from unauthorised access, manipulation, and hacking. Due to the sensitive nature of IoT data, it is essential to implement strong security measures, including authentication, encryption, access control, and device management.

The data may encompass financial, personal, or commercial information. Ensuring user privacy and compliance with data protection regulations, such as the General Data Protection Regulation (GDPR) and the Consumer Protection Act, are essential for the widespread usage of the Internet of Things (IoT). The seventh aspect is data lifecycle management, which involves the management of data across its entire existence. Effectively managing the whole data lifecycle of the Internet of Things is crucial for maximising its advantages and minimising its risks. Collecting, using, manipulating, examining, and disposing of data are all components of this process. In order to effectively handle IoT data while ensuring security, compliance, and cost-effectiveness, it is necessary to have a scalable infrastructure, data lifecycle management procedures, and efficient data management software. Data is the driving force behind the functionality, insights, and value generation of the Internet of Things (IoT). To fully utilise linked devices, individuals and organisations must comprehend the attributes and difficulties associated with IoT data in the current age of technology.

1.2 CHARACTERISTICS OF 'BIG DATA'

Big data, as the name suggests, pertains to vast quantities of data and datasets that are characterised by three main attributes: size, speed, and variety. The datasets come from a variety of sources, including the Internet of Things, sensors, and posts on social media. The administration and storage of big data are constantly evolving in response to the growing use of contemporary technology by a larger population. Businesses may gain significant benefits by collecting and evaluating data for user insights and industry trends, rather than discarding this valuable data resource. The ability of businesses to effectively navigate the data visualisation process is where big data truly brings value, as opposed to the unprocessed data itself. However, the intricate nature and large amount of unorganised and organised data render conventional methods of information processing and analysis inadequate for handling these invaluable collections.

Thankfully, recent developments in artificial intelligence (AI) and machine learning (ML) have made it feasible to turn vast amounts of data analytics into insightful and helpful knowledge. This facilitates companies in making informed decisions.

The term "Internet of Things" describes the integration of sensors and actuators into physical objects, which are subsequently networked together using wired and wireless connections. Frequently, these networks make use of the same Internet Protocol (IP) infrastructure as the internet as a whole. This concept envisions a world where a multitude of diverse entities are interconnected over the internet and capable of exchanging information.

Nevertheless, the precise meaning of "Big Data" remains fairly complicated. While volume, diversity, and velocity are commonly associated with it, other attributes might also be significant, especially in operational contexts. The fundamental components of "Big Data," which are typically represented by the recognizable Vs, are shown in Figure 1.2 and further described in Table 1.1. Essentially, "Big Data" refers to the vast quantities of well-organized, partially-structured, and unstructured data that consistently inundate businesses. Volume, Velocity, and Variety are the three key factors that characterize it. However, throughout time, more attributes have emerged that offer a more intricate perspective on large data.

The term encompasses the wide-ranging, diverse, and constantly evolving nature of data in several domains, encompassing aspects such as expansion, intricacy, and availability. These attributes often surpass the computational capabilities of traditional database systems and software programmers.

Big data are datasets that are larger than what can be processed by standard data processing methods. Big data management demands the use of sophisticated tools, techniques, and technologies, including data mining, Hadoop, and Spark platforms, in addition to classification algorithms, since it surpasses the capabilities of traditional analytical systems. Big data is often emphasized for its significant size, variety, and speed, leading to the exploration of new approaches for managing, processing, and evaluating data.

"Big data" in the context of smart, sustainable cities usually refers to massive volumes of urban data that pose significant logistical and computational challenges due to their enormous scale and the need for real-time analysis. The majority of urban big data comes from automated processes and sensors, and it's frequently annotated with geographic and temporal information.

Although there isn't a single, agreed definition for big data, it's generally agreed that it will have a big impact on a lot of industries in the future, accelerating development and creating new opportunities. Along with other characteristics like truth and usefulness, big data is sometimes defined by four essential characteristics, also known as the Vs: volume, variety, velocity, and. These features draw attention to the large scale, diversity, and speed at which big data analytics processes data. "Big data analytics" refers to the systematic collection, archiving, processing, and analysis of massive datasets to extract crucial information needed to make decisions. Big data analytics utilizes state-of-the-art software and high-performance computers in the context of smart sustainable cities to provide valuable insights that can be applied in many urban domains, such as energy, transportation, and the environment. Various urban challenges are tackled using a range of analytical methods, including descriptive and predictive techniques, frequently including machine learning and statistical analysis.

Urban analytics is a field that applies data science principles to extract valuable insights from urban data in order to tackle sustainability challenges. Regression analysis and other approaches, such as data mining, are crucial for comprehending urban dynamics and guiding decision-making. For those who are interested in learning more about urban analytics and how it contributes to urban sustainability, Bibri (2018) is a useful resource.

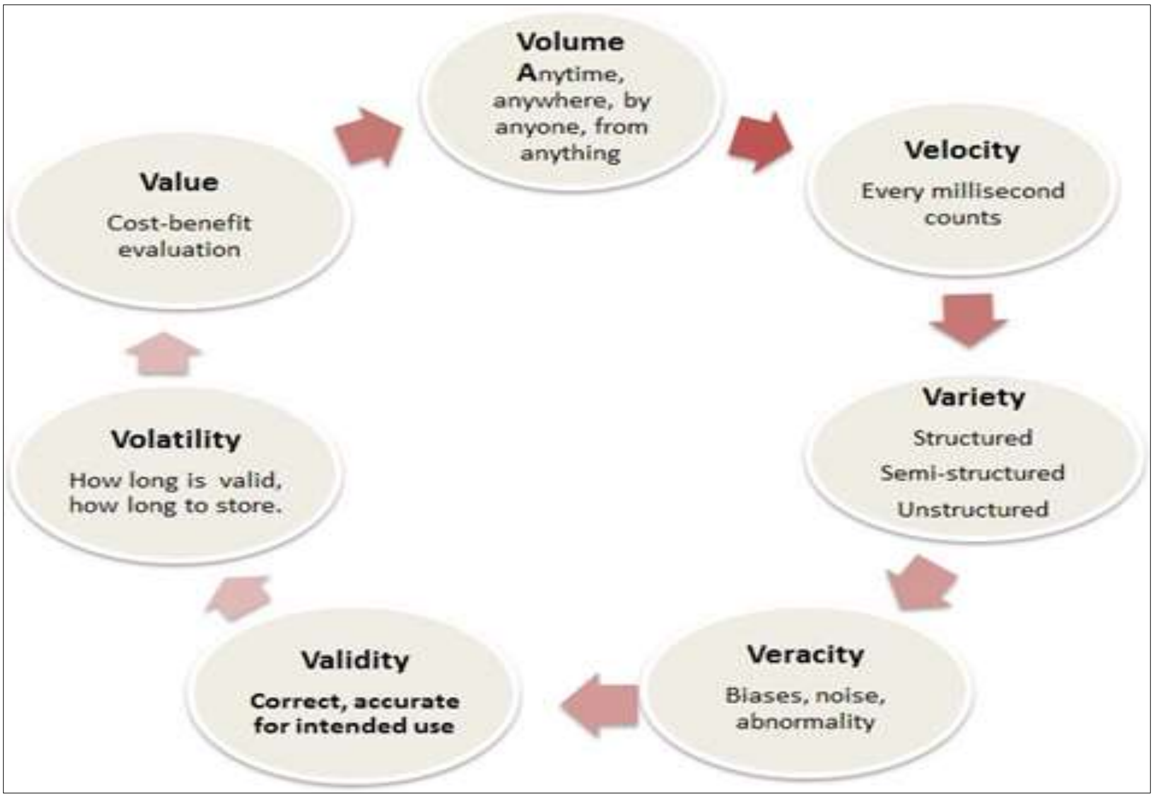

Figure 1.2: The Definition of Big Data.

Table 1.1: The Characteristics of the Data in IoT.

Attributes	Description	Instance
Quantity	Currently available applications do more than just log business transactions, move samples, and compile them into another database; they now gather all available data for analysis. These applications collect consumption data from various sensors attached to various appliances and devices; readings are taken from each home every day at a rate of roughly 100,000 times.	Anywhere, at any moment, and by anything
Speed	Modern apps gather streaming data from external systems or sensors, where the data is continuously produced.	Every millisecond matter
Variety	There are several forms of data that can be obtained, such as unstructured, semi-structured, and structured forms. 1. Computers can comprehend structured data easily and it has semantic meaning.	Database data
	2. A type of structured data known as semi-structured data deviates from the formal structure that is usually connected to data models in relational databases.	XML and additional markup languages
	3. Unstructured data is so devoid of intrinsic meaning that a computer is unable to understand how it is represented.	The Emails, text messages, audios, and videos streams
Veracity	Tackling biases, noise, and anomalies within data presents a more significant hurdle compared to concerns about volume and velocity. The implementation of data cleaning procedures and safeguards against the influence of "dirty data" are critical components in tackling this challenge.	dirty data
Validity	Validating data becomes imperative if a specific segment holds significance or is deemed important.	correct and accurate data
erratic	In certain cases, analysts can create retention guidelines and policies for big data by having a thorough understanding of the data that is currently available and how long it might be stored. This does not, however, hold true in situations where analysis of the data is consistently possible.	The duration of the data's validity and storage.
Value	Because the implementation of "Big Data" requires infrastructure for data collection, storage, and processing, this "value" plays a critical role in determining the cost-benefit analysis of the use of Big Data.	cost-benefit evaluation

1.3 THE DATASET COLLECTION & DESCRIPTION

In the smart grid system, the various components are employed to gather data, with the smart meter serving as the primary device for data collection from homes, buildings, and other designated locations. Despite the necessity for a substantial volume of smart meter data for analysis, researchers encountered a shortage of collected datasets. Consequently, for one of the projects in Baghdad, the research opted to utilize electrical energy consumption data. The study concentrated on how much energy was used overall each day in particular households. The dataset utilized comprises readings from 3,567 Baghdad households participating in the Baghdad Low Carbons Electricity Networks projects, spanning from November 2021 to February 2023. These readings encompass different data points available for each half-hour interval, enabling analysis of average daily consumption on an hourly basis. However, relying solely on this dataset proves inadequate. To enhance analysis and yield optimal results, additional features need to be incorporated. Weather data, known to significantly impact energy consumption, was integrated into the dataset. Daily weather information was obtained using the Dark Sky API, providing a date-based platform for comprehensive data analysis. Hence, the research relies on a combined dataset comprising meter data and weather information to facilitate robust analysis and yield meaningful insights.

2. TOPICAL LITERATURE REVIEW

This approach reviews all of the prior research on the subject, including new findings in big data analytics, the fundamental technologies supporting the IoT and big data analytics, and any relevant applications, challenges, or gaps in our understanding. The study delves into significant philosophical and scientific obstacles specific to IoT and big data analytics, in addition to providing a concise summary of outstanding issues. In order to perform this literature review, a wide array of materials from different disciplines that are related to the study subject will be evaluated, with a main concentration on qualitative research. These sources may include books, journal articles, dissertations, conference proceedings, and reports. To identify the relevant components, build search algorithms for discovering articles and documents across many sources, and design inclusion and exclusion criteria for locating suitable information, a systematic review procedure was developed and applied.

Many academic studies on big data have been made available as a consequence of the research on the literature of intellectual output undertaken in the databases and sources of publisher information made available through the Egyptian Knowledge Bank. One such study (Wang & Ruan, 2016) may shed light on the influence of big data in a new era of library transformation and conclude that smart services for libraries She came to the conclusion that libraries must continue to construct and maintain their systems in order to utilise big data and improve their smart services. Research (Akbar, 2017) indicated that the amount of data being created is significant and requires treatment strategies at the national level. It also revealed that the advantages of big data are still limited when compared to the untapped potential and values. Furthermore, the research concluded that legislation relevant to information privacy and intellectual property should be put in place.

Additionally, Hanson & Golub (2017) talked about how libraries have historically benefitted from their concentration on gathering, organising, preserving, and giving access to information sources for diverse user groups in communications. They also mentioned the availability of big data in the context of quantitative scientific expansion, organisation, and data sharing—three data-related areas in library and information science that were examined in this paper from the standpoint of achieving the significance of the beauty of research and the study discovered that library and information science is unusual in that new technologies, Big Data, and networked research environments are continually rising in number and diversity. Through educational activities and the introduction of new research disciplines like data processing and analytics, it tries to take advantage of rising prospects. Widén and Zhan's 2017 study sought to develop a comprehensive understanding of big data. Given libraries' unique role in big data management and utilisation, the study concentrated on understanding big data in libraries according to the definitions used in that field. This helped to clarify the current understanding of big data in libraries. A compilation of articles containing definitions of 35 definitions was made because there were few definitions analysed in the study and the content and sta.

The scale, capacity, and diversity of data are the first of two major issues facing academic libraries when applying big data analytics, according to Karoui-Al & Barashdi-Al (2018, Karoui-Al & Barashdi-Al). The second difficulty is based on: Thus, this study sought to address how librarians engage with big data, future research directions, and gaps in big data studies pertaining to academic offices. The researchers also carried out an extensive literature review on big data analysis. The goal of this study was to investigate the methods and resources that can be used in academic libraries to analyse big data and then decide how to allocate them to improve the economics of academic libraries. the findings led to the publication of 37 big data-related publications in academic libraries, and the findings showed that despite the volume of the effects of big data on academic libraries, including analytical methods and tools, have not, however, been extensively explored in research.

In the same context, al-Omari's (2018) study sought to determine the reality of big data in academic libraries in the Sultanate of Oman. It used a quantitative descriptive approach, and its sample consisted of 106 library specialists from 28 academic libraries. The study's findings showed that, among the sources of big data collected in academic libraries, searches for electronic sources yielded the highest percentage (69%) of results. Meanwhile, predicting future library needs emerged as one of the most notable uses of big data, reaching 75.5%, indicating that librarians' ignorance of the significance of big data and its analysis came in last. This number rose to 71.4% when it was discovered that the two biggest obstacles to big data are the high expense and limited availability of technical skills (68.9%).

A study(al-ahammateya et al., 2018)has confirmed the importance of statistical data as an essential pillar in decision-making and strategic planning, and it also forms a large part of big data, as it is linked with other data to give it added value, and these data may often remain useless if it is not provided with analysis and good management of their preservation, retrieval and availability, and the study the National Center for Statistics and information in the Sultanate of Oman represents a vital role in the task of providing and supporting various sectors and institutions through the provision of statistical data Due to the importance of statistical data, this study aimed to explore the reality of statistical data production, identify the nature of its availability, the means used in publishing it, and ways to invest it, as well as identify the factors affecting the production of statistical data at the National Center for Statistics and information in the Sultanate of Oman. By presenting the previous models of intellectual production literature, it is noted that they were characterized by focusing on big data from the perspective of concepts and importance and how to use them to provide the model of information services targeted by libraries, especially how academic libraries participate in big data, while the current study is working on addressing the stages of work in the management of big data better content personalization and improved user experience.

Big data analytics plays a crucial role as a foundational technology for realizing the Internet of Things (IoT) within the realm of pervasive computing. It involves extracting valuable insights from vast datasets to enhance decision-making across various urban domains. In recent years, the scope of big data analytics has expanded beyond traditional business intelligence, encompassing urban development areas such as energy, transportation, and environmental sustainability. This evolution has positioned big data analytics as a key component of the ICT infrastructure of smart sustainable cities, particularly within the IoT ecosystem. In this context, big data analytics aims to optimize and support intelligent decision-making processes related to urban system control, automation, and management, ultimately enhancing ecosystem services and environmental sustainability. Despite its potential, literature on the application of big data analytics for environmental sustainability in smart sustainable cities remains limited, reflecting the early stages of research in this emerging field.

Various reviews and surveys have been conducted on big data analytics, offering diverse perspectives and highlighting different computational and analytical aspects. However, many of these studies are oriented towards business intelligence and focus on specific components of big data analytics such as techniques, algorithms, and software tools. To address the lack of survey work specifically focused on the core enabling technologies of big data analytics within the context of smart sustainable cities, Bibri and Krogstie (2017c) conducted a comprehensive review of the technical literature. Their aim was to identify and synthesize these technologies and illustrate key computational and analytical techniques associated with the functioning of the big data ecosystem in smart sustainable cities.

Bibri and Krogstie proposed conceptual frameworks to guide further research in this area, providing a deeper understanding of how big data analytics can contribute to urban sustainability. These frameworks are intended to add depth and rigor to studies in the field and stimulate further research endeavors. They cover essential elements such as data collection and preprocessing, data repositories, processing techniques, analysis algorithms, wireless network technologies, and data visualization. Additionally, the authors highlight key applications of big data analytics for urban sustainability, though further analysis is needed to understand their functioning within the context of smart sustainable cities and environmental sustainability. Thus, there is a need for thematic analysis and the development of analytical frameworks to advance research in this area.

3. DATA MANAGEMENT AND APPLICATIONS

In the context of the Internet of Things, there are several ways to look at the underlying conceptual and physical structure of data management. IoT data management has four main processes: data administration, processing, data transmission over available communication channels, and data collection. These procedures are designed to simplify the management of information technology data. The primary goal of the data gathering module is to identify "entities" that can provide data to IoT data repositories. Prior to notifying the appropriate regulatory authorities, this data may be retained for a prolonged duration. After undergoing processing, filtering, and, in certain circumstances, fusion into condensed forms, data is effectively sent from the bottom layer of "things" to the top layer.

Wireless communication technologies, such as Zigbee, cellular networks' application layer, and Wi-Fi, can be employed to transport data to collecting locations. The process of analysing data is facilitated by the use of middleware and a database design that incorporates essential elements such as query functionalities and access control. Key components involve the optimisation and aggregation of data. Following this, we will outline some stages that demonstrate the significance of implementing a proficient data management system for the Internet of Things. Table 3.1 provides a concise overview of pertinent studies on IoT data management, specifically emphasising database and data gathering technologies.

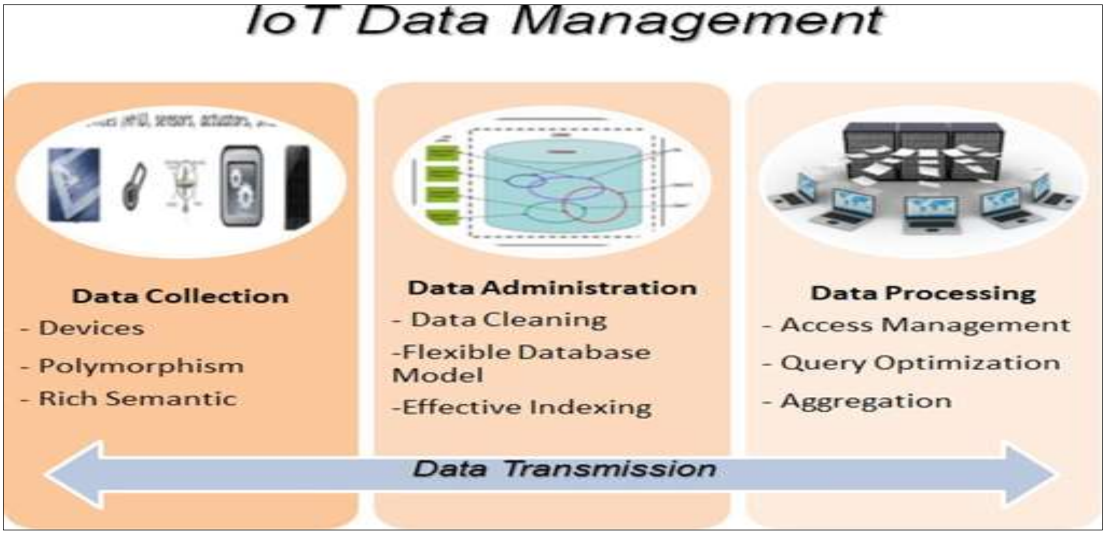

Figure 3.1: The Basic Concept IoT Data Management.

Table 3.1: The Use of Databases, Data Types, And Sources in The Internet Of Things.

Ref	System	Data Source							Database		Application Area
		RFID	Smartphone	Sensor	GPS	NFC/WSN	PDA	Actuator	SQL	NoSQL	
34	Effective index structure utilizing a key-value store (UQE-Index)				P					P	HBase
23	Processing sales and inventory using OLAP DBMS (parallel database)	P							P		MySQL SQL Server
30	MDB system for IoT (NoSQL)			P						P	Mongo DB
25	Mobility First's routing based on identity			P	P			P	P		SQLite
31	document-focused for multimodal and heterogeneous data				P					P	CouchDB
32	Cluster for General Statistics Databases (IoT-Statistic DB)			P	P				P		PostgreSQL
61	NoSQL for the WSN population			P						P	Redis
62	XML Resource Directory Powered by Database					P			P		MySQL, Sedna
33	The Discovery Service (DS)					P				P	Hbase

3.1 HOW BIG DATA AND THE IOT WORK TOGETHER

Although the Internet of Things (IoT) and big data are distinct concepts, they synergize in synthesizing and processing information for the benefit of businesses. While IoT serves as a primary data source for big data, the latter comprises vast amounts of information gathered from diverse sources, including IoT networks.

IoT data predominantly consists of real-time data derived from customer interactions with connected devices, offering richer and more immediate insights compared to other sources. However, this data is raw and the sheer volume generated by IoT user devices can overwhelm traditional analytical methods without appropriate tools and processes.

To fully leverage the potential of IoT data and derive valuable insights, businesses must undertake a structured process for data extraction. Typically, IoT big data processing involves several key steps:

a. Data Collection: Automated interactions among Internet-connected physical devices gather extensive data from customers using software products over a specific timeframe.

b. Data Storage: The collected data is aggregated into files within a big data database and organized according to business requirements.

c. Data Analysis: Analytical systems are employed to automate the analysis of this vast dataset.

d. Reporting: Enterprises categorize the analyzed data into descriptive (real-time device performance), diagnostic (understanding device output), predictive (forecasting device behavior), and prescriptive (recommendations for impacting predicted outcomes) insights. Reports generated from these insights inform software product development, tuning needs, and decision-making processes.

3.2 BENEFITS OF THE IOT AND BIG DATA RELATIONSHIP

a. Enhance customer experience through real-time data analysis: Numerous factors influence a business owner's decision-making process, with users holding paramount importance as the ultimate purpose of a software product is to cater to their needs. User needs are dynamic and evolve over time, highlighting the significant power of real-time data products facilitated by IoT networks.

Real-time insights offer a snapshot of users' current state and capture any significant events occurring during their product experience. By promptly collecting and analyzing data pertaining to user habits at the moment of data generation, businesses gain insight into customer preferences and pain points, enabling them to implement corresponding improvements to enhance the overall user experience.

Moreover, beyond mere enhancements, big data harvested from expansive IoT networks, commonly observed on e-commerce platforms and social media sites, facilitates the delivery of personalized content, advertisements, and recommendations in real time, leveraging user browsing history.

b. Forecast the Most Current Trends: As the latest trends continually evolve, people's lives become increasingly diverse. Being the first to identify upcoming trends in their respective fields can elevate a business to prominence in the eyes of customers. Recognizing the importance of trend anticipation, businesses leverage the synergies of Internet of Things (IoT) and big data to forecast and spearhead trends, thereby attracting users.

In particular, big data analytics employ machine learning algorithms to anticipate future trends by analyzing historical data collected from various connected devices. By scrutinizing past trends and patterns, businesses can make informed predictions about the next wave of trends.

Moreover, real-time data from IoT networks provides valuable insights for businesses to swiftly respond to emerging trends. For instance, sensors monitoring environmental conditions can alert businesses to potential issues such as high humidity or impending thunderstorms. This enables businesses to proactively communicate with customers, providing early warnings and mitigating potential problems.

Despite numerous benefits, IoT combined with big data analytics presents a host of challenges. Data collected from connected devices are often perceived as factual representations of products and users. However, the complexity of big data can lead to deception and render it valueless to businesses if analyzed incorrectly.

Therefore, data scientists must exercise caution in classifying and extracting data, identifying errors, duplicate information, and other anomalies to ensure accurate insights and seamless data tracking. Avoid excessive data collection and focus solely on relevant information by identifying reliable data sources and narrowing the scope to meet business needs.

Incorporating IoT and big data into business operations can be daunting, particularly for startups. At Orient Software, we comprehend the challenges and aspirations of our clients, drawing on decades of experience in the technology industry. Beyond offering a diverse range of services, we also provide consultancy to guide you through the software product development process. Contact our experts for tailored advice and support.

3.3 COMPONENTS OF THE INTERNET OF THINGS (IOT):

The "Internet of Things" refers to the network connectivity and processing capabilities given to ordinary objects, sensors, and devices that are not often seen as computers. Thanks to this add-on, these devices are now capable of autonomously generating, distributing, and using data with minimum human interaction. The term "Internet of Things" (IoT) refers to a network of interconnected computing devices that enable smooth transmission and exchange of data. Key components include:

a) Devices, objects, and things are tangible items in the physical world that have the ability to collect and send data through their built-in sensors, actuators, and communication capabilities.

b) Sensors and Actuators: Devices can acquire data about their environment using sensors, and subsequently execute physical actions in response to that data using actuators.

c) Connectivity, thirdly, refers to the mechanism via which devices establish connections with one other and with the broader internet using different communication networks. Mobile phone networks, Bluetooth, wireless networks, and Zigbee are among the most often utilised communication systems.

d) Data Processing: The Internet of Things (IoT) produces vast quantities of data. Data processing involves the examination and interpretation of data, which can be done either on the device itself or in a remote server known as the cloud.

e) Furthermore, cloud computing is important for the storage, analysis, and administration of the immense quantities of data produced by Internet of Things (IoT) devices.

f) The sixth component is middleware, which acts as an intermediary software layer that enables efficient communication and seamless data exchange between IoT devices and the applications or services they interact with.

g) Edge Computing: Occasionally, data processing can be performed in proximity to the origin, at the "edge" of the network, rather than only depending on centralised cloud servers. This enhances real-time processing and minimises delay.

h) Security: The Internet of Things (IoT) offers protection against possible dangers and unauthorised entry to data and devices.

i) Apps and Services: These refer to the software applications or user interfaces that utilise the data and functionalities provided by IoT devices in order to give unique insights or services.

The smooth functioning of communication and compatibility across various Internet of Things (IoT) platforms and devices relies on the implementation of standardized protocols. Protocols such as MQTT, HTTP, and CoAP are widely used.

These components, when integrated into the Internet of Things ecosystem, provide a network of networked devices that facilitate the gathering, processing, and utilization of data across various services and applications.

a. Foundational Concepts: The research starts by examining the fundamental principles behind Big Data and the Internet of Things (IoT). This section explores the essential attributes of big data, such as its size, speed, variety, and integrity. This text explores the core aspects of IoT ecosystems, such as sensors, actuators, and communication protocols, in greater depth.

To investigate the basic ideas behind Big Data management within the Internet of Things (IoT) framework, a thorough literature review is necessary. This type of analysis must consider the following factors.

b. IoT BDA and challenges of relational database management systems (RDBMS): Although data analytics has a long history, the rapid increase of data created by IoT devices presents substantial difficulties for traditional systems used for collecting, storing, managing, and analysing data, such as RDBMS.

These systems are not specifically intended to handle the overwhelming amount of data generated by Internet of Things (IoT) devices in smart environments. This data is primarily unstructured and relies on events. This difficulty is present in a wide range of geographies and sectors.

The convergence of IoT and BDA can impact several areas of RDBMS, including flexibility, real-time responsiveness, anomaly detection, integration of advanced hardware and software, and compatibility with other systems. Based on the research findings, in order to enhance traditional systems for Internet of Things (IoT) Big Data Analytics (BDA), the following enhancements are necessary:

a) Flexible data platforms: Cheng et al. (2015) identified stiff data platforms as the main hindrance to the implementation of IoT BDA. These outdated systems are incapable of provide the necessary insight and context for apps and services. Adaptability is essential for the successful functioning of IoT BDA. In order to handle the increasing number of data sources, future analytics systems must possess the capability to maintain high performance, predict future requirements, and work together with other tools to explore, collect, store, and oversee large volumes of data. Informatica, Datameer, and Ayasdi are viable alternatives for adaptable IoT corporate data analysis platforms.

According to Howell et al. (2017), the integration of BIM with the smart water industry will remove operational obstacles. In a similar vein, Shapsough et al. (2016) challenge the prevailing notion of smart education by asserting that the Internet of Things surpasses the limitations of traditional forms of online education, such as massive open online courses (MOOCs). Tan et al. (2016) argue that traditional techniques of analysis and data management are unable to meet the requirements of modern enterprises. They contend that organisations must adapt in order to simultaneously enhance staff productivity, customer happiness, and product quality and reliability.

b) Next generation software-hardware solutions: In their 2015 study, Hans et al. attacked traditional parking systems, referring to them as "outdated" due to their inability to meet the requirements of smart cities that utilise parking sensors for monitoring and controlling the system. These sensors enhance the efficiency of parking operations and ensure smooth traffic flow.

A Real-Time Operating System (RTOS), as proposed by Venkataraman and Chitra (2015), together with additional software solutions such as cloud storage, mobile applications, a PHP web gateway with a MySQL database, and Python, as described by Hans et al. (2015), are important for the advancement of future smart systems (2015).

c) Real-time response systems (RTRS): One of the drawbacks of relational database management systems (RDBMS) is their sluggish response time when confronted with unforeseen circumstances. In their study, Sheriff et al. (2015) propose that the implementation of complex event processing (CEP) and comparable technologies might enhance the ability of next-generation emergency systems to accurately identify the origins of incidents and promptly inform consumers.

The Internet of Things has greatly advanced real-time response systems (RTRS). The main goal of the Internet of Things is to enable the creation of smart environments through the effective interchange of real-time data among independent and interconnected components. The automatic updating of emergency scenario status, demands, and other relevant information can enable prompt response through the use of RFID sensor networks and other technologies such as GPS or infrared sensor detection.

Rizwan et al. (2016) stresses the need of real-time systems in smart city settings for rapidly updating traffic statistics. They contend that the existing traffic management and alarm systems may be substituted by these affordable, real-time technologies, as the current systems lack the necessary sophistication to effectively control traffic.

3.4 BIG DATA TOOLS AND INDUSTRY SUPPORT

The Industrial Internet of Things (IIoT) has experienced exponential increase in data generation due to significant breakthroughs in sensing, networking, technical characterization, and high-throughput instrumentation [9]. Lin et al. [10] suggested a method for integrating data from several sources and devices in order to enhance Industrial Internet of Things (IIoT) applications. It was discovered that IIoT can mitigate problems related to database transmission and various protocols in industrial data testing. This paper presents a comprehensive multimedia big data (MMBD) classification system that is then used to develop a new process model that incorporates MMBD within the Internet of Things (IoT).

It explores the intricacies of MMBD computing for IoT applications. The process model, which covers many MMBD research issues such service quality needs, heterogeneity, dependability, accessibility, and scalability, is demonstrated in a case study [11]. Furthermore, a proposal is made for flood monitoring and prediction that utilises the integration of large-scale data with high-performance computing (HPC). This structure enables the storage, analysis, gathering, and prediction of flood consequences [12].

The integration of mobile computing services into the Internet of Things can be accomplished by utilising mobile devices, apps, or M-Healthcare systems [13]. Alexopoulos et al. [14] provided a comprehensive explanation of the structure and development details of the Industrial Internet of Things (IIoT) to facilitate the life cycle of the industrial product service system. Furthermore, in the realm of industrial growth, a novel framework is formulated to analyse the fundamental components of big data analytics (BDA). Businesses may utilise the findings to gain a deeper understanding of the possible ramifications of big data and to create infrastructures that enable big data analytics for their industrial operations [15]. Boyes et al. [16] were the pioneers in introducing the concept of Industrial Internet of Things (IIoT) and its connection to concepts such as cyber-physical systems and Industry 4.0. With a focus on security issues and vulnerabilities, they created an analytical framework for IIoT devices and looked at Internet of Things taxonomies.

We provide and execute a system for efficient choice selection and analysis of sentiment in large datasets (BDSA) utilising mathematical methodologies [17]. Prior research has shown a novel architecture for CIoT (Cellular Internet of Things) and large-scale data [18]. The system utilises both data lakes (DL) and data warehouses (DWH) together with a selected approach for gathering various data. Urquhart and McAuley [19] discussed the regulatory and technological factors related to the hazards associated with IIoT. Furthermore, an examination was conducted on the structural and functional characteristics of cloud manufacturing (CMfg), and a business intelligence framework was suggested for the dissemination of pertinent key performance indicators (KPIs) associated with process data [20].

We conducted a product life cycle analysis and examined the use of big data in smart manufacturing. Additionally, we put up a recommended application design. This approach enables the analysis of primary advantages, possible uses, and exploration of future areas of study and existing obstacles [21]. This research investigates current methodologies, tools, and algorithms used in big data analytics (BDA) to enhance intelligent IIoT systems. Scientific classifications consider important elements such kinds of analytics, applications in industry, needs, methodology, tools, and sources of data. Frameworks and case studies from various firms that have utilised BDA may be found [22]. Furthermore, this study examines how businesses might utilise big data to improve environmental engagement by conducting an extensive analysis of relevant literature that establishes connections between big data sources and the implementation of sustainable initiatives [23]. Big data and the internet of things (IoT) have several more uses apart from these tactics to address diverse real-world situations [24–28].

3.5 BIG DATA – HOW TO MEET IOT.

The integration of IoT and Big Data plays a pivotal role in various aspects of our lives. IoT generates a substantial volume of data crucial for Big Data analytics, which is efficiently captured, processed, and analyzed. This synergy allows companies to gain a comprehensive understanding of their business, customers, products, and competitors, leading to potential improvements in efficiency, increased sales, cost reduction, enhanced customer service, and improved products and services.

The term "big data" refers to the utilization of larger data volumes for visualizing scientific data, giving rise to the concept of the "Internet of Signs." This notion suggests that data generated on the internet, encompassing IoT devices, social media, and other online sources often associated with Big Data, forms "signs" reflecting sentiments toward various issues. This aggregation of "signs" creates an "Internet of Signs," offering insights into potential events and situations.

Big Data and IoT collectively have the power to drive implementation, working alongside other software and hardware components to realize the vision of smart cities. These cities are envisioned to bring transformative changes, addressing issues such as pollution, waste, parking problems, and enhancing energy savings. The practical applications include:

a. Measuring and regulating traffic with RFID tags on cars, sending geo-location data to a central monitoring unit to identify congested areas.
b. Providing citizens with real-time information on public transportation status and availability through smartphones and mobile devices.
c. Enhancing child safety in parks through sensor-equipped bracelets that allow tracking in case of a child going missing.
d. Utilizing Big Data to reduce emissions and pollution, with road sensors measuring traffic and emissions at different times of the day. The data can be sent to a central unit coordinating with traffic police for effective management (K.R.Kundhavai et al, 2016).

3.6 BIG DATA ANALYTICS

In his 2015 publication, Mike Butler defines Big Data analytics as the utilization of powerful tools and techniques by data scientists, analysts, and statisticians to unveil trends and patterns within vast, unstructured, and dissimilar datasets. The goal is to make these insights readily accessible to business leaders, managers, and key stakeholders, informing the development of business strategies and plans. The field encompasses four types of data analysis:

a. Descriptive Analytics: This type focuses on answering the question "what happened." It involves reviewing data using traditional research approaches and often employs Classical or Bayesian statistical methods. An example would be analyzing the average monthly amount of money in a bank account.

b. Diagnostic Analytics: Diagnostic analytics aims to answer "why did something happen." For instance, it delves into understanding why a customer chose to purchase a DVD player from a specific company when alternatives were available on the shelf and online.

c. Predictive Analytics: Predictive analytics seeks to answer "what will happen in the future." This branch is dedicated to forecasting actions or events. An example includes generating a sales forecast for a business.

d. Prescriptive Analytics: Prescriptive analytics endeavors to determine "what should be done." An example is a stock portfolio optimization model that not only predicts potential outcomes but also suggests how the portfolio should be allocated accordingly.

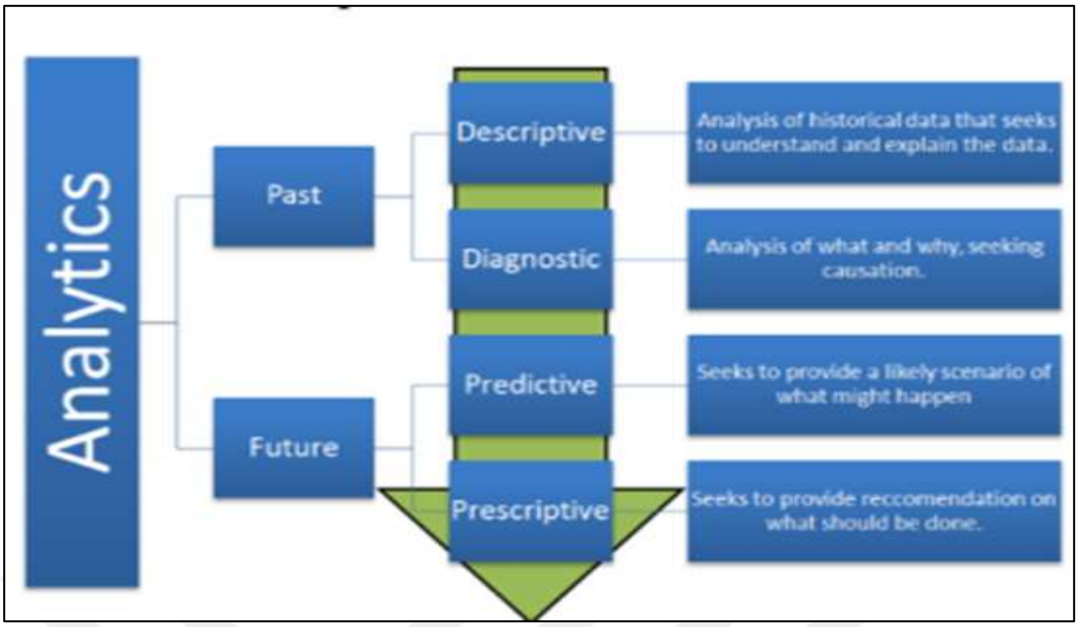

Figure 3.2: Big Data Analytics.

3.7 CLOUD COMPUTING FOR BIG DATA ANALYTICS

Cloud computing is a word that has several definitions due to the varying interpretations of ICT specialists, academia, and corporations. Cloud computing is often defined as a systematic approach to providing standardised, scalable, adaptable, and real-time information and communication technology (ICT) over the Internet. Usually, these skills are classified into three primary service categories: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). Infrastructure as a service (IaaS) encompasses the virtual servers, storage facilities, processors, and networks offered by a provider and housed on the cloud. Software as a service (SaaS) and platform as a service (PaaS) refer to the software applications and development platforms provided by a service provider. This strategy enables rapid and effortless allocation of resources, with minimal involvement from management. However, some individuals have contended that the existing definitions of "cloud computing" are excessively expansive and vague, resulting in a dearth of agreement over the exact essence of the idea.

Individuals from all backgrounds and governmental entities at all levels can utilise the diverse functionalities offered by cloud computing, such as database administration, data exchange and storage, and the development of online services. Another use of cloud computing is the processing of large datasets, which is especially beneficial in scientific research and for developing smart and sustainable cities.

Cloud computing has gained significant popularity and widespread use in recent times, particularly in the context of smart cities. The rise can be attributed to the progress made in sensing devices, data storage, computer platforms, widespread computing infrastructures, and wireless communication networks. The growing appeal of cloud computing can be attributed to the advanced state of these technologies and their cost-effectiveness. Cloud computing tackles various challenges encountered by smart sustainable cities by commoditizing services, offering affordable (and sometimes free) software, providing extensive data storage, and supplying tools essential for data processing, analysis, and administration. Cloud computing facilitates a range of big data applications in the context of intelligent and sustainable cities, while also providing the necessary infrastructure for dispersed urban computing. To meet the needs of urban residents, powerful computers must be deployed in large data centres spread throughout different locations. More precisely, it involves promoting urban sustainability by utilising tools, approaches, strategies, models, and technology for analysing large sets of data.

Cloud computing, which was introduced in 2011, involves the utilisation of virtualized computer processing resources that are dynamically accessible over the cloud. Therefore, in metropolitan settings, the only physical goods required are display gadgets for information and services. Administrators, planners, landscape architects, and occupants, together with other stakeholders, can access software applications and services using web browsers, lean clients, or mobile devices. This access is possible even if the software tools and urban data are kept remotely on servers. Smart, sustainable city applications are kept in cloud-based storage and may be accessed at any moment when needed. From an ecological standpoint, these challenges encompass transportation, electricity generation, traffic congestion, and waste disposal. Furthermore, clients are relieved of the burden of configuring and sustaining specialist platforms as the cloud provider assumes responsibility for the software development platform and offers it through public, private, or hybrid networks.

Cloud computing offers several benefits, such as enhanced service operations, decreased expenses for information and communication technology, and reduced energy use. Cloud computing reduces electricity usage and greenhouse gas emissions by efficiently using processing power and distributing resources among several users. Furthermore, it employs techniques like as energy-aware scheduling and server consolidation to facilitate energy conservation.

Moreover, cloud computing powered by sustainable energy sources such as solar panels is regarded as a type of environmentally friendly computing. It facilitates the pooling of resources and reduces costs in ICT operations. It utilises supercomputers located in huge data centres to provide scalable services and effectively handle enormous amounts of data.

Cities have a significant need for computer capacity when utilising vast data analytics to inform their choices on environmental sustainability. Cloud computing offers a financially efficient answer to this issue. Through the utilisation of cloud computing, smart sustainable cities may improve their efficiency and effectiveness in all areas of urban planning and administration. The use of cloud computing also enables service-oriented computing. Urban residents can benefit from speedier processing of enormous amounts of urban data, which improves environmental sustainability performance.

3.8 THE PROBLEMS OF USING BIG DATA IN INFORMATION ORGANIZATIONS AND WAYS TO OVERCOME THEM

Al-Akbar, 2017, reviewed many of the challenges facing information organizations, perhaps the most prominent of which are the fear of their inability to absorb and control the continuous flow of big data, as well as challenges related to their traditional systems used in data mining and analysis, in light of the rapid growth of data and The Associated problems of processing and storage space, search and retrieval issues, in addition to threats to the privacy of big data, etc. Cloud storage systems represent one of the proposed solutions to overcome data storage and processing issues, however, this solution is not accepted by some organizations, especially in the case of sensitive data that is subject to confidentiality requirements, and fear of being breached, especially with the emergence of problems related to information security, and this study suggests to overcome such a problem the privatization of a it adopts strict procedures to maintain privacy and secure data and information.

Information organizations can resort to the provision of specialized techniques in efficient data management, the model of which is the use of Advanced Dynamic Data Processing Techniques and data density calculation and management. In order to be able to overcome the obstacles of search and retrieval, information institutions should adopt mechanisms leading to the infrastructure of the data warehouse with sufficient capacity, data analysis, characterization and organization, as well as employing the capabilities and capabilities of international search engines to enhance search and retrieval capabilities, which is one of the requirements of Big Data Management, and the potential and requirements of using administrative and organizational skills, monitoring and acquisition of big data, experience in dealing with big data, knowledge of big data organization tools and skills of using them. On the other hand, new risks of system penetration may arise due to the possibility of accessing a large amount of data, and it was noted (2016, al et Wang) that research and studies of the use of big data in libraries did not take into account the issues of data security in the field of data protection and the use of large amounts of data in the IT has not been given sufficient consideration and attention. The current study should note that there is a sector that believes that the sharing and dissemination of users ' personal data through social networks may not fall within the general concept of privacy, and this sector is based on its vision that big data platforms aim, through the analysis, management and storage of this data, to increase its value and provide specialized and effective services accessible to users, and therefore, in this context, the breach of private Big Data. After this focused presentation on the issues and limitations facing the use of big data in information organizations and addressing the mechanisms for overcoming them, the following paragraphs examine the role of the Hadoop platform in managing big data.

3.9 BIG DATA AND ITS ROLE IN THE SPECIALIZATION OF CONTENT AND SERVICES

Big Data technologies can play an important role in helping information organizations to customize and enhance content directed to users, and it is established that customization requires, first of all, obtaining and collecting as much personal data of users as possible, analyzing "Logs Search "search records, user behaviors and orientations systematically and applying machine learning mechanisms, in order to identify user preferences and extract content that contributes to services effectively. Institutions ,information centers and academic digital repositories of universities, as well as entities dealing with content from a commercial perspective, represented by publishers and suppliers of database services, including, for example, Elsevier - Ebsco - Proquest, are characterized by their continuous work to increase the number of visitors and subscribers in their sources and services, and (2020, Nelson)shed light on the main problem that it lies in how these entities can make use of the main amount of sources to improve search and personalization. Big data systems represent a major tool that information organizations can rely on to maximize opportunities to benefit and invest in their sources and enhance their services, as these organizations can achieve successes in the specialization of content and services using big data by: - collecting primary data about the user and content and the source of this data can be the personal files of the beneficiaries, search queries, downloaded files, keywords and beneficiaries, metadata and taxonomy taxonomy preparation and others. - Processing and analysis of user log data and content within the Hadoop system. - Feeding search engines with content, data characterization and other mechanisms that enable them to provide relevant results and unique recommendations through the browser interface, whether it is a website, a smartphone application, etc. - Through the "Learning Machine Automated "machine learning, powerful cross analytics can be performed between users' browsing habits (unstructured data) and content (structured data) to provide a personalized search and browsing experience. - Personal data such as: current location, study qualification, general and exact specialization, age, gender, date of initial contact, etc.from all available data that are allowed to be collected, and if the user works in the information organization, this can include any employee data such as: years of experience, Department, unit and Administrative Department affiliated with him, job title, etc. - Registration of the activities carried out by these users on the portal of the information institution, such as: the pages viewed, the sections visited and the use of their content and services. - Activities on other websites and web pages such as: clicks that the user makes on advertisements, referral links, etc. - Find out the user's membership in thematic interest groups, if any, and identify the interests and activities of these groups, for example: if the user is known to be a member of an institution, Association, or office group.

The activities that the user performs in external sources such as: social networks and others. Activities carried out through other systems and applications. - Bribery activities. The study proposes through Figure 3.3 the general infrastructure through which big data can be used for the purposes of personalizing the content and services of Information institutions, and this structure consists of the information institution portal on the global network, a distributed database that includes events and events, a database distributed with the features and interests of users and beneficiaries of the sources and services of the information institution, and the search and retrieval database that complies with the standard "XML".

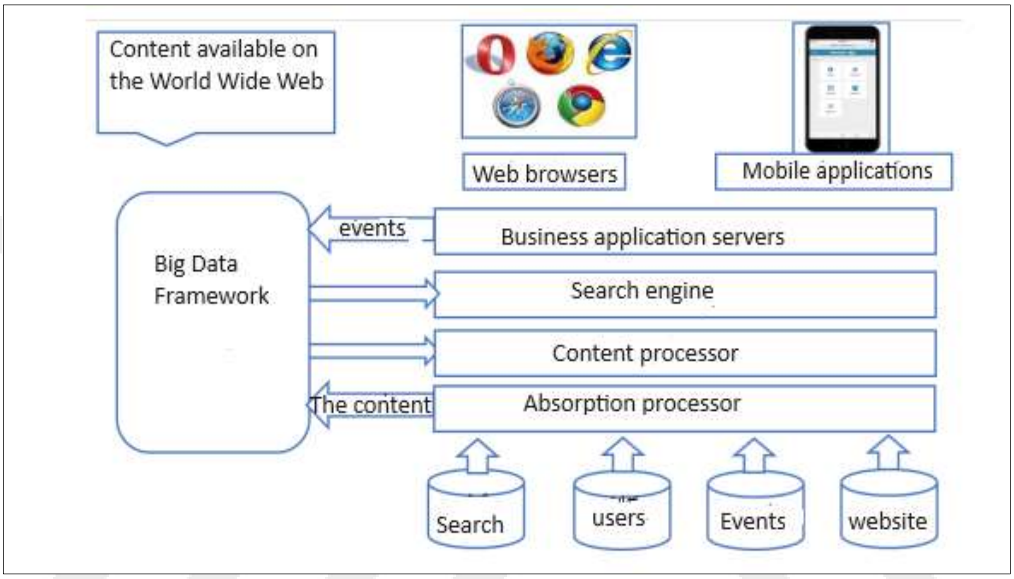

Figure 3.3: Big Data Architecture for Customizing Content and Services.

The previous components are associated with a set of layers, which are represented in the data absorption, input and collection layer and work to process.

These inputs are "processing injection", then comes the content processing layer "processing Content", followed by the layer responsible for performing , And the "application Business search" factor is represented by your engine search engine or search tool in the application server layer servers", a layer that deals directly with user requests, whether through mobile applications and / or browsers To make use of the content and services in a special way. These layers communicate directly with the Big Data Framework, which receives inputs represented by unstructured data After managing and processing such data, it is sent to both the content processing layer and the search engine،

It is worth mentioning that the Big Data Framework is connected to the content and sources that can be accessed via the global network.

4. THE PROPOSED SYSTEM

In our research, we have reached a point of conviction in taking a model that connects all of our words mentioned in our paragraphs above, which is the idea of collecting big data and linking it to the Internet of Things, and this idea is collecting data about the amount of electrical energy spent in the seasons and according to the air temperature, relying and relying on that according to new studies that we explain in Our following paragraphs show the results we obtained according to k-means algorithms and equations.

The stages that make up the system follow a distinct sequence and are interdependent. An electricity metre, similar to a gadget, is utilised to quantify the current instantaneous electrical energy usage. Subsequently, it transmits this data to a specifically assigned storage facility over the Internet. Subsequently, the consumption data is merged with the latest weather forecasts. Subsequently, data processing generates outcomes through the process of gathering, evaluating, and identifying significant attributes. Subsequent to the preprocessing stage, there will be a discussion on these results. The conceptual foundation for these phases of the system is depicted and elaborated in Figure 4.1 below. The research will utilise the following excerpts.

4.1 K-MEANS

The processes that make up the system follow a distinct sequence, and each step is dependent on the others. An electricity metre, similar to a gadget, is utilised to quantify the current instantaneous electrical energy usage. Subsequently, it transmits this data to a specifically assigned storage location across the Internet. Subsequently, the consumption data is merged with the latest weather forecasts. Subsequently, data processing generates outcomes by including the activities of gathering, evaluating, and choosing significant qualities. Subsequent to the preprocessing stage, an analysis and interpretation of these data will be presented. The conceptual framework for these stages of the system is depicted and elaborated in Figure 4.1 below. For the purpose of our investigation, we will make use of the paragraphs that follow.

Though NP-hard in terms of computational complexity, effective heuristic methods rapidly approach a local optimum. The expectation-maximization strategy used in Gaussian mixture modeling is comparable to these tactics. Both methods iteratively enhance cluster assignments. Although both k-means and Gaussian mixture models utilise cluster centres for data modelling, k-means is more adept at identifying clusters with comparable geographical extents, whereas Gaussian mixture models are capable of handling clusters with varied forms.

The popular k-nearest neighbor classifier, which uses supervised machine learning for classification, has a weak correlation with the unsupervised k-means technique. The names' resemblance may cause misunderstanding. The 1-nearest neighbor classifier is applied to the cluster centers produced by the k-means technique in the Rocchio method, also referred to as the closest centroid classifier. This approach categorises incoming data into existing clusters.

The name "k-means" was coined by James MacQueen in 1967, but the notion was originally proposed by Hugo Steinhaus in 1956.Pulse-code modulation was initially developed in 1957 by Stuart Lloyd at Bell Laboratories, although it was not officially published in a scientific publication until 1982.The user's text is "[4]". The approach is commonly referred to as the Lloyd-Forgy algorithm due to Edward W. Forgy's independent publication of a similar method in 1965.

4.2 PREPROCESSING DATA

Preprocessing techniques involve converting raw data into a format that is easily understandable. This vital phase, referred to as data preprocessing, is an essential component of data mining. It entails preparing data for the mining operation by performing operations like cleaning and reshaping it. The goals of data preparation include reducing data volume, discovering relationships within the data, standardizing it, removing outliers, and extracting pertinent features.

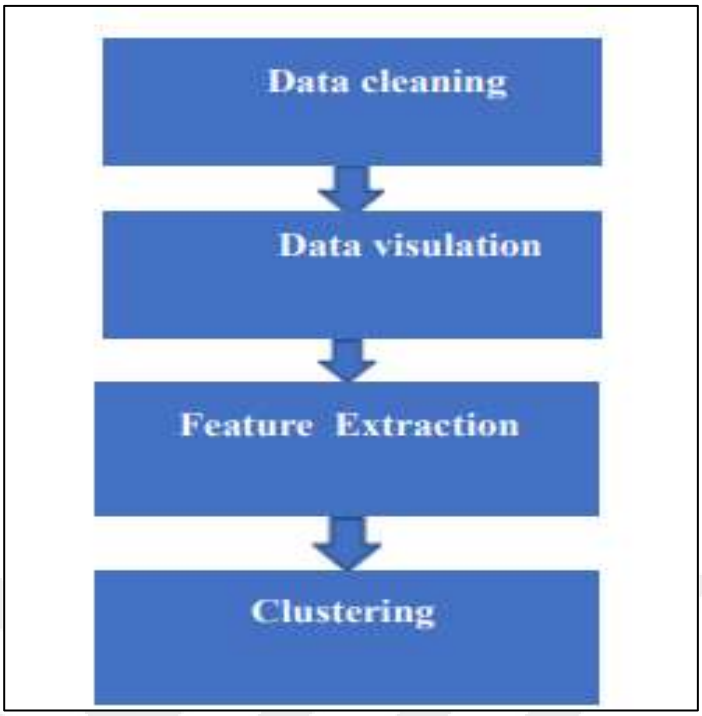

Figure 4.1: The Data Preprocessing.

This stage encompasses various techniques such as data cleaning, integration, transformation, and reduction [23]. It serves as a crucial preparatory phase preceding the application of machine learning or data mining methods. Its primary goal is to guarantee the data's quality and format suitability for the intended analysis.

DATA CLEANING

Data cleaning entails eliminating duplicate, damaged, inaccurate, poorly structured, or incomplete data from a dataset. When merging multiple data sources, duplication or mislabeling can occur in various ways. While categorizing the specific procedures within the data cleaning process isn't precise, methods will vary from one dataset to another. Nevertheless, establishing a model for the data cleaning process is crucial to ensure consistency and accuracy [24]. Although the techniques for cleaning data may differ based on the types of data stored, research typically adheres to a series of steps, including removing redundant or irrelevant observations, correcting structural errors, filtering out outliers, addressing missing data, and conducting validation and quality assurance checks.

4.3 DATA VISUALIZATION

Visualization plays a crucial role in today's corporate landscape, which heavily relies on data. It enhances the understanding of complex data by including elements such as colour, form, location, and height. Data scientists frequently utilize graphical representations such as graphs, charts, and maps to evaluate the outcomes of investigations. Data visualization offers a comprehensive view, leading to its extensive use in enhancing decision-making processes [25]. Choosing suitable data visualizations is a crucial component of every data visualization strategy. Multiple visualization techniques are accessible, as seen in Figure 4.2. The visualizations available include heat maps, area plots, scatter plots, line plots, pie charts, graphs, and corruptibility maps.

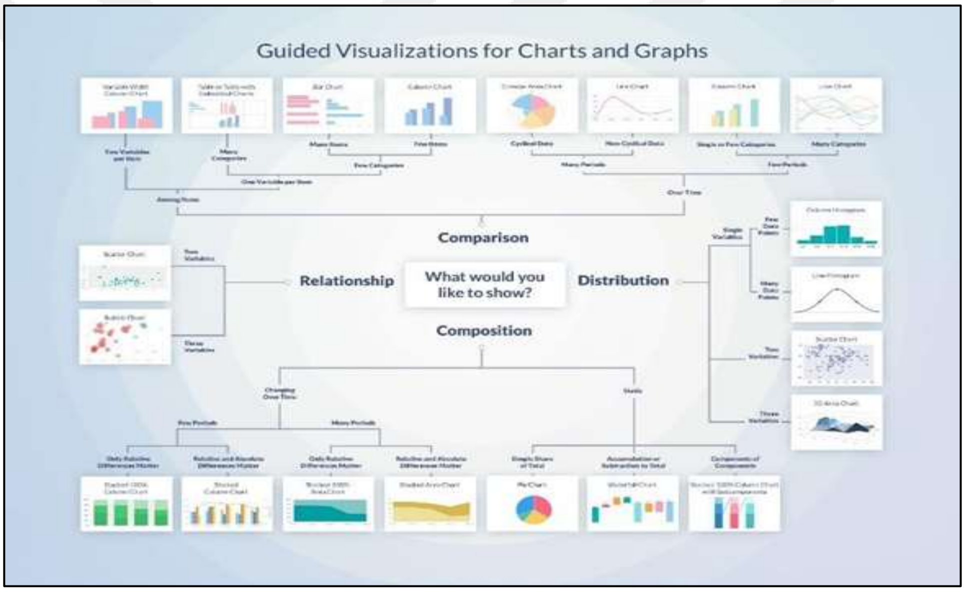

Figure 4.2: Visualization Graphs Type.

The temporal nature of the data on energy consumption led to the use of time series graphics. A time series comprises data points collected over time intervals, enabling the analysis of past behavior and prediction of future trends using various statistical methods. This method is predicated on the notion that success values are recorded on a regular basis [26]. Time series analysis has two main goals and covers a wide range of phenomena: Understanding the series' properties and forecasting values for future trends using data from the past and present. Achieving these objectives necessitates the formal definition and description of a sequential model [27].

4.4 FEATURE SELECTION

Upon receiving a dataset, numerous features are often uncovered. However, not all features are beneficial for constructing a machine learning model for predictive tasks. Indiscriminate utilization of all features can potentially degrade model performance. Hence, the selection of appropriate features is paramount in model creation. While there's a desire for a model with optimal predictive power and interpretability, achieving both simultaneously can be challenging, often necessitating a trade-off between the two. Feature selection enhances model interpretability while simplifying its complexity. Limiting the model to only the most pertinent features enhance interpretability and aids in constructing practical models for the studied phenomena [28]. Feature selection techniques in machine learning fall into the following categories [29]:

a. Supervised methods: These approaches are useful for identifying relevant features that enhance the performance of supervised models in labeled datasets. They are applied to categorical data.

b. Unsupervised Techniques: These methods are adept at analyzing unlabeled data. When it comes to classification approaches, a variety of feature selection algorithms have been proposed. The majority of these algorithms utilize statistical metrics such as mutual information, correlation, and information gain measures [30]. There are three main strategies for feature selection based on evaluation metrics: Filter, Wrapper, and Embedded methods.

4.5 THE CORRELATION COEFFICIENT

Correlation coefficient formulas are used to assess the degree of relationship between two features within a dataset, as illustrated in Figure 4.3. The results of these computations range from -1 to +1 [31], where the:

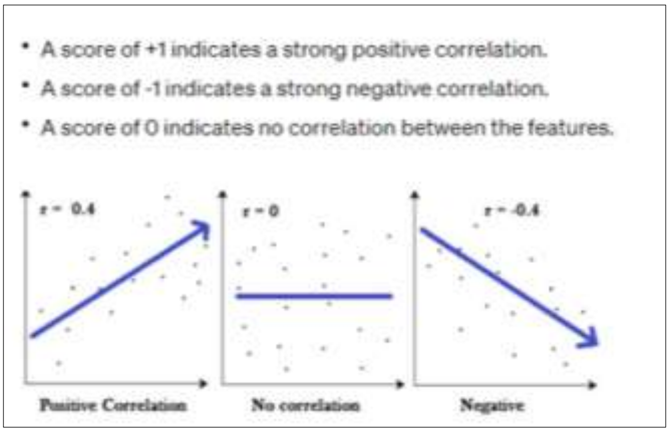

Figure 4.3: Correlation Coefficient Relationships.

The correlation coefficient's magnitude indicates the strength of the relationship between two features [32].

$$r = \frac{\sum_{i=1}^{n}(xi-\bar{x})(yi-\bar{y})}{\sqrt{\sum_{i=1}^{n}(Xi-\bar{X})^2 \sum_{i=1}^{n}(yi-\bar{y})^2}} \tag{4.1}$$

5. THE IMPLEMENTATION AND RESULTS

5.1 CLEANING

Data cleaning include replacing or eliminating inaccurate data as well as noise, duplicate entries, and blank fields. Additionally, you'll need to fill in the blanks with your typical daily consumption rate. The next step is to calculate the daily average power consumption of each building by taking the total consumption of each building and dividing it by the total number of structures. In Figure 5.1, the x-coordinate displays the days, and the y-coordinate indicates the number of meters recorded on each day. This data is used to calculate the total number of meters collected.

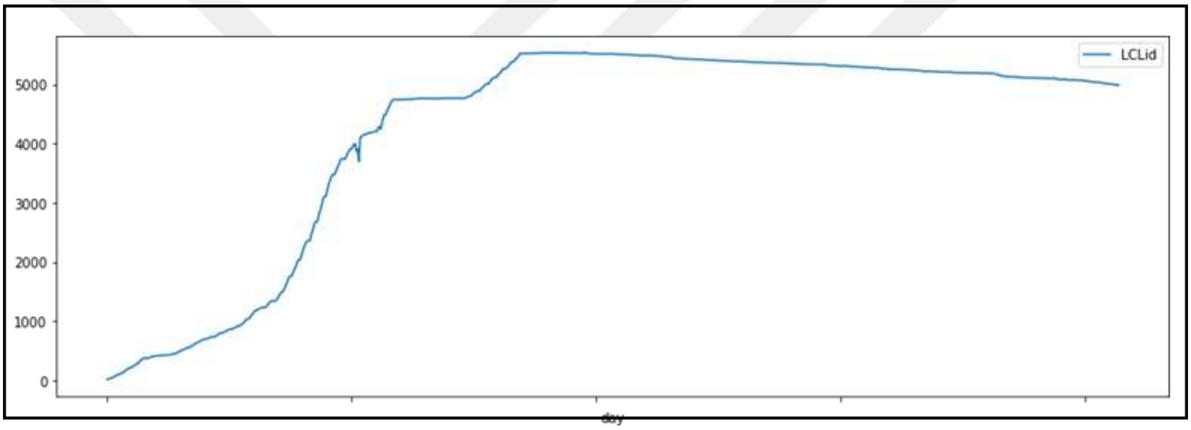

Figure 5.1: Counts the Meter per of the Days.

5.2 VISUALIZATION

The goal of data visualization is to make it easier to understand analytics results by presenting a comprehensive overview of the data using visual components such as charts, graphs, and maps. When looking for correlations between weather and energy consumption over a certain time frame, it really shines. By visualizing the effect of different weather conditions on electricity consumption, we may have a better understanding of how much each factor affects consumption percentages.

5.3 ELECTRICITY'S THE USAGE COMPARED TO TEMPERATURES

Temperature stands out as a pivotal weather factor that notably impacts energy consumption, as illustrated in Figure 5.2 below.

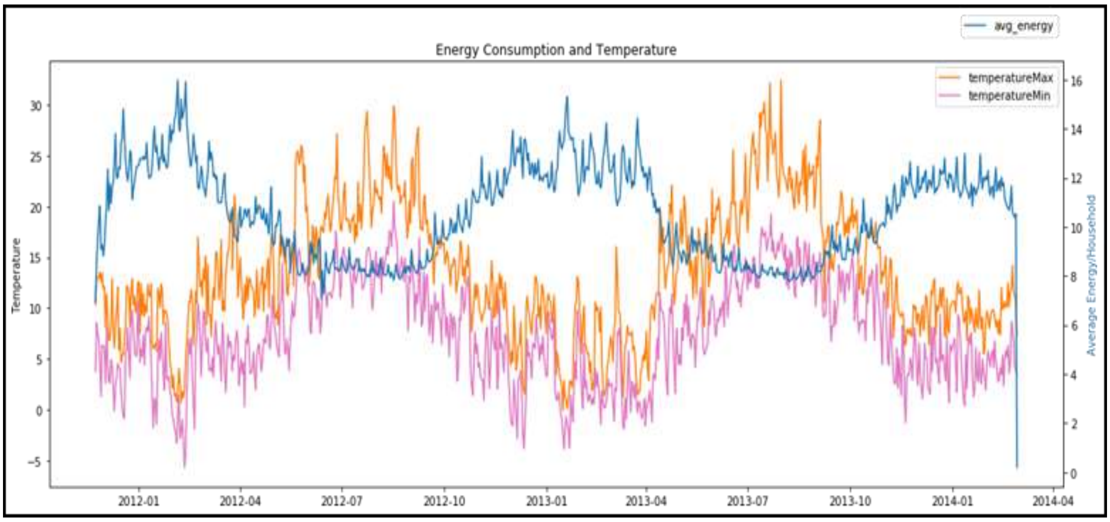

Figure 5.2: Correlation Between Energy and Temperature.

An inverse relationship between temperature and energy use is suggested by the blue peaks, which represent electricity consumption, contrasted with the dips in the orange (highest temperature) and pink (lowest temperature) regions. This study lends credence to the commonly accepted notion that, as the temperature drops, energy consumption tends to rise, perhaps because people turn on their heaters more often to stay warm.

5.4 HUMIDITY

Similar to temperature, humidity is a weather variable that doesn't significantly impact energy use. Because the correlation between humidity and average energy usage does not show any obvious trend, it can be concluded that there is no discernible association between the two. Energy consumption in a humid environment remains constant independent of the humidity level, as seen by the continuous orientation of the colours in Figure 5.3 below:

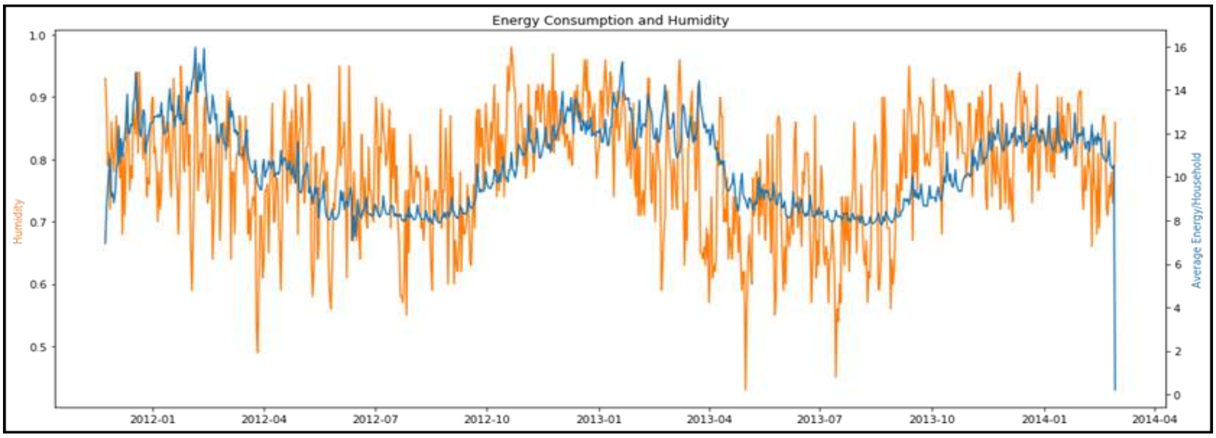

Figure 5.3: The Relationship between Energy and Humidity.

5.5 CLOUD COVER

Cloud cover is another weather factor that does not exert a significant impact on the energy consumption, as depicted in Figure 5.4 below:

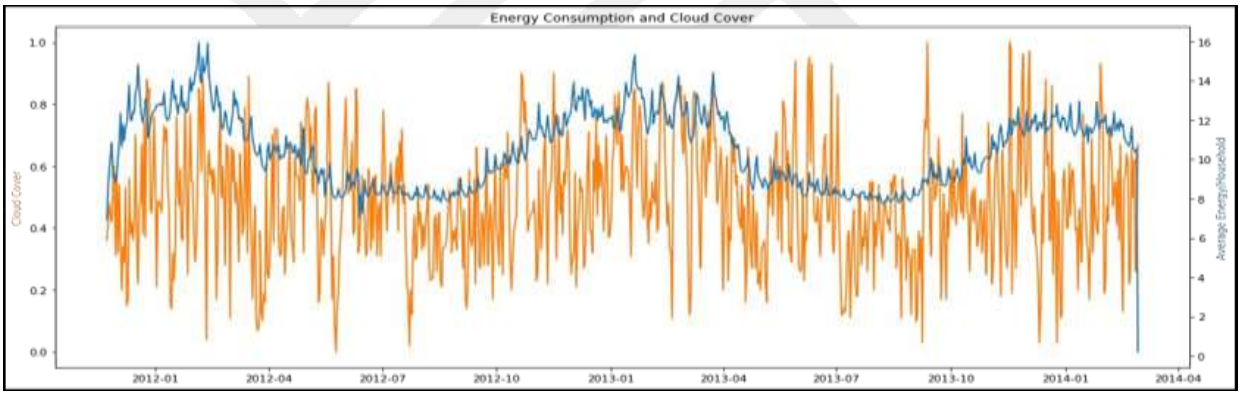

Figure 5.4: The Relationship between Energy and Cloud Cover.

There seems to be no relationship between changes in cloud cover and changes in energy consumption; instead, variations in cloud cover and changes in humidity seem to follow the same pattern.

5.6 FEATURE SELECTION

A thorough comprehension of the weather factors influencing power use was achieved once the data visualisation phase was finished. To identify the most significant features with the greatest impact ratios, a correlation matrix was employed. The correlations between weather and power use are illustrated in Figure 5.5, which is located below:

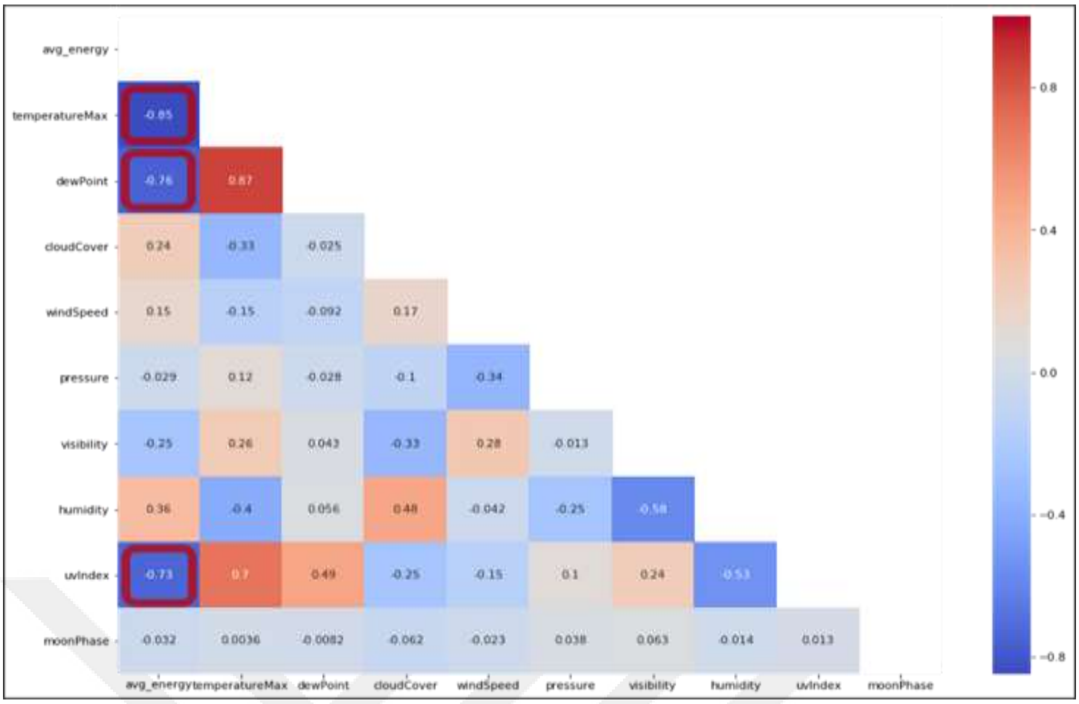

Figure 5.5: Presentation of Correlation Matrix.

The correlation matrix presents correlation coefficients between different variables, including meteorological factors like cloud cover, wind speed, pressure, humidity, UV index, visibility, energy consumption rate, maximum temperature, dew point, and moon phase. The matrix's rows and columns each show the correlation between every possible combination of these variables. The strength of the relationship between intersecting factors is indicated by values within the matrix cells, which range from +1 to -1. While the sign denotes the type of relationship (positive sign denoting a positive relationship, negative sign denoting an inverse relationship), the absolute value indicates the magnitude of the relationship. This analysis identifies three features: temperature, dew point, and UV index, whose absolute values are greater than 0.6.

5.7 WEATHER AND DATA CLUSTERING

After selecting the strong weather features, five clusters of these features are created using K-Means. This division of data into clusters is essential for subsequent steps, as weather conditions are required, and utilizing clusters provides a more efficient way to determine weather conditions. Additionally, when applying certain artificial intelligence techniques in subsequent steps, it is preferable to use summarized data for faster processing. Figure 5.6

illustrates the relationships between the weather factors (temperatures, dew points, UV index) selected in preciouses step and the resulting weather clusters.

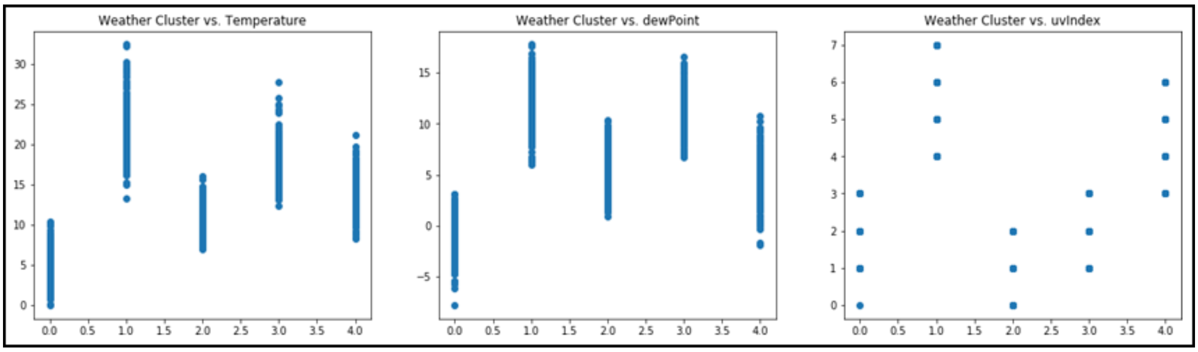

Figure 5.6: Weather Cluster.

Figure 5.7 provides an alternative visualization of the relationship between the weather cluster and selected features. Each dot in the figure corresponds to a day, and four parameters are represented by each dot: the X-axis denotes temperature, the Y-axis denotes dew point, the dot size indicates the UV index, and the dot color signifies the cluster class.

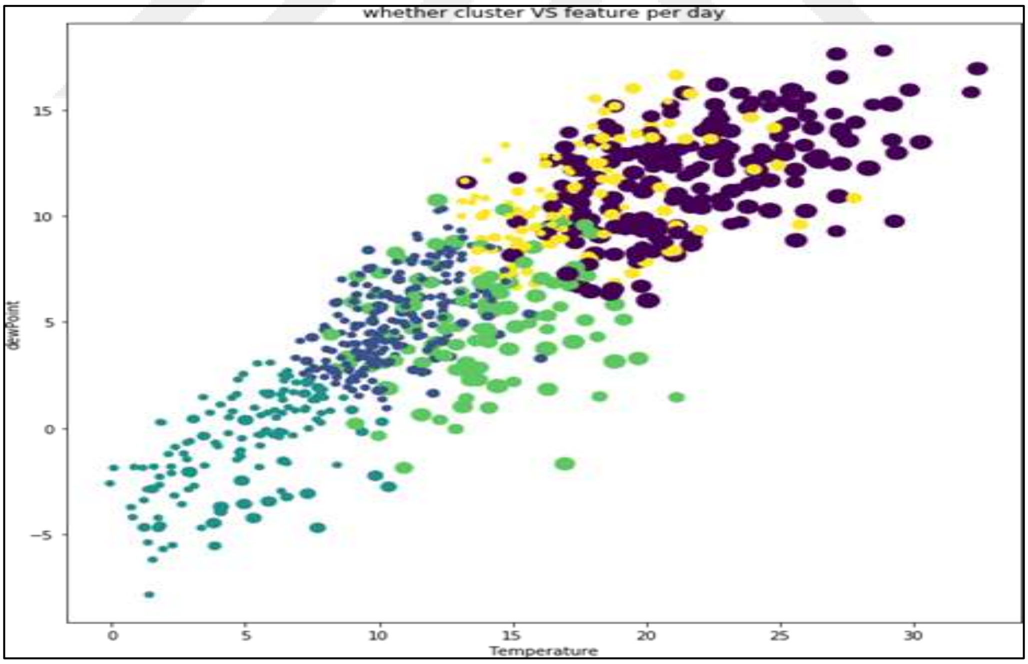

Figure 5.7: Daily Clustering of Electricity Consumption and Weather Data.

6. CONCLUSIONS

This paper describes how to collect, analyze, and display energy consumption data from internet-connected electrical meters in relation to temperature, dew point, UV index, and other meteorological variables. The main goal is to gather data on energy usage from buildings in a certain area to determine the overall consumption for that region. Data cleaning, visualization, feature extraction, and clustering are all components of the preparation step that follows. Data cleaning entails the elimination of flawed, replicated, inaccurate, or unfinished entries. The process of turning abstract data into insightful knowledge is called visualization. Finding important characteristics is the process of feature extraction. Clustering is the process of dividing data into groups according to certain characteristics; the k-means algorithm is used to achieve this goal.

The study examines the relationships between the selected climatic factors that have a major impact on energy use. Five distinct groups are identified from the K-Means clustering of meteorological data; these groups each reflect a consistent relationship between temperature, dew point, and UV index. In future research, these weather clusters and daily electrical consumption data will be used to estimate the energy usage of the region.

REFERENCES

[1] EIOPA, Gert. "Artificial intelligence governance principle: Towards ethical and trustworthyartificial intelligence in the European insurance sector." (2021).

[2] EIOPA, Gert, "Big Data Analytics in Motor and Health Insurance: A Thematic Review (European Insurance and Occupational Pensions Authority)"., (2019).

[3] Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. "Artificial intelligence: the ambiguous labor market impact of automating prediction." *Journal of Economic Perspectives* 33.2 (2019): 31-50.

[4] Jang, Jeong Ah, Hyun Suk Kim, and Han Byeog Cho. "Smart roadside system for driver assistance and safety warnings: Framework and applications." *Sensors* 11.8 (2011).

[5] Masters, Jon. "Telematics making fresh gains." *ITS International* 18 (2012).

[6] Shannon, Darren, et al. "Connected and autonomous vehicle injury loss events: potential risk and actuarial considerations for primary insurers." *Risk Management and Insurance Review*, issue. 24.1, P.P 5-35., 2021

[7] Titsworth, Tammy. "Telematics might steer your car into the future." *IEEE multimedia* 9.3 (2002): 9-11.

[8] Cieślik, Barbara. "Telematics in automobile insurance." *Collegium of Economic Analysis Annals* 45 (2017): 79-92.

[9] Constantinescu, Cornel Coca, Ion Stancu, and Iulian Panait. "Impact study of telematics auto insurance." *Rev. Financ. Stud*. No. 3. P.P.17-35., 2018

[10] Naseri, Raghda Awad Shaban, Ayça Kurnaz, and Hameed Mutlag Farhan. "Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach." Applied Soft Computing 134 (2023).

[11] Husnjak, Siniša, et al. "Telematics system in usage based motor insurance." *Procedia Engineering* 100 (2015): 816-825.

[12] Yoon, Daesub, et al. "Future automotive insurance system based on telematics technology." *10thInternational Conference on Advanced Communication Technology*. Vol. 1. IEEE, 2008.

[13] Cunneen, M. A. R. T. I. N., and M. A. R. T. I. N. Mullins. "Framing Risk, The New Phenomenon of Data Surveillance and Data Monetisation; from an 'Always-On'culture to 'Always-On'artificial Intelligence Assistants." *In Proceedings of the International Conference on Robot Ethics and Standards, New York, NY, USA*. 2018.

[14] Cunneen, Martin, Martin Mullins, and Finbarr Murphy. "Artificial intelligence assistants and risk: framing a connectivity risk narrative." *Ai & Society* 35 (2020): 625-634.

[15] Al-Obadi, Mohamed Ghanim, et al. "Data mining techniques for extraction and analysis of covid-19 data." *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*. IEEE, 2022.

[16] Cunneen, Martin, et al. "Autonomous vehicles and avoiding the trolley (dilemma): vehicle perception, classification, and the challenges of framing decision ethics." *Cybernetics and Systems* 51.1 (2020): 59-80.

[17] van den Boom, Freyja. "Regulating telematics insurance." *Van den Boom, F (2020) Regulating Telematics Insurance, in P. Marano and K. Noussia (Eds): Insurance Distribution Directive, AIDA Europe on Insurance Law and regulation* 3 (2020).

[18] Irode, Philip L. *An Auto Telematics System for Insurance Premium Rating & Pricing*. Diss. University of Nairobi, 2017.

[19] Al-Aloosi, Ahmed Raad, et al. "Face recognition system using local binary pattern with binary dragonfly algorithm to feature selection." 2022 International Conference on Artificial Intelligence of Things (ICAIoT). IEEE, 2022.

[20] Kuo, Kevin, and Daniel Lupton. "Towards explainability of machine learning models in insurance pricing." *arXiv preprint arXiv:2003.10674* (2020).

[21] Ma, Yu-Luen, et al. "The use of context-sensitive insurance telematics data in auto insurance rate making." *Transportation Research Part A: Policy and Practice* 113 (2018): 243-258.

[22] Khalef, Dhirgham Atia, et al. "Optic disc segmentation in human retina images with meta heuristic optimization."International Conference on Artificial Intelligence of Things (ICAIoT). IEEE, 2022.

[23] Filipova-Neumann, Lilia, and Peter Welzel. "Reducing asymmetric information in insurance markets: Cars with black boxes." *Telematics and Informatics* 27.4 (2010): 394-403.

[24] Kargupta, Hillol. "Connected cars: How distributed data mining is changing the next generation of vehicle telematics products." *Sensor Systems and Software: Third International ICST ConferencecLisbon, Portugal, June 4-5, 2012,*

[25] Cunneen, Martin, et al. "Artificial driving intelligence and moral agency: Examining the decision ontology of unavoidable road traffic accidents through the prism of the trolley dilemma." *Applied artificial intelligence* 33.3, P.P 267-293, 2019).

[26] Vallor, Shannon. *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press, 2016.

[27] Mizgier, Kamil J., Otto Kocsis, and Stephan M. Wagner. "Zurich Insurance uses data analytics to leverage the BI insurance proposition." *Interfaces*, No. 48.2, P.P 94-107, 2018.