Chapter 9
Big Data Applications in Smart Sustainable Energy Systems and E-Mobility: Review and Case Study

Tayyibah Khanam, Mohammad Saad Alam, Sanchari Deb, and Yasser Rafat

Abstract Since sustainability-related concerns need to be answered beforehand, a vast majority of research in the recent past has been focussing on the potential utilization of data-driven approaches to move a step ahead in achieving sustainable development goals. While big data in itself is not technological advancement, practices such as analysis of data sets and implementation of Machine Learning algorithms on data sets can give deeper insights and understanding to aid the decision-making process of organizations. With the increasing amount of data in almost every sector, it is now even possible to get accurate predictions for practical purposes. Hence, big data sciences can be viewed as a paradigm shift in achieving sustainable development goals. First-hand, this chapter gives necessary background about sustainable development, renewable energy, big data, and E-mobility. Since smart cities are expected to integrate several important components of human–nature interactions in the near future, we perform a review of the existing literature to examine the current research directions in the role of big data tools and technologies in different components of smart cities. Additionally, pieces of literature on the applications of big data in electric vehicles are closely reviewed for better understanding & problem formulation. Lastly, this chapter also presents a case study based on traffic congestion data to analyse and understand the charging behaviour of users across the city of Mumbai and thus draw relevant conclusions regarding the Charging Pile Network of the city.

Keywords Big data · Charging infrastructure · Renewable energy · Smart cities · Sustainability

T. Khanam (✉) · M. S. Alam
Department of Electrical Engineering, Aligarh Muslim University, Aligarh, India
e-mail: tayyibahkhanam@zhcet.ac.in

M. S. Alam
e-mail: saad.alam@zhcet.ac.in

S. Deb
Electric Power Train Group, VTT Technical Research Centre, Espoo, Finland

Y. Rafat
Department of Mechanical Engineering, Aligarh Muslim University, Aligarh, India
e-mail: yasser.rafat@zhcet.ac.in

I. Ali et al. (eds.), *Computational Modelling in Industry 4.0*,
https://doi.org/10.1007/978-981-16-7723-6_9

1 Introduction

The requirement for sustainability emerges from the right to extremely essential needs like food, water, energy, sanitation, and work to everybody. Absence of these essential needs brings about poverty. Till date, almost half of the world's total populace, i.e. 3 billion people live in poverty and about 297,000 children under five kick the bucket every year because of destitution (11 Facts About Global Poverty 2021). Apart from poverty issues, another significant danger to life on earth is air contamination due to the consumption of fossil fuels. These fossil fuels discharge GHGs (green-house gases) and other harmful gases which are an immediate danger to an unnatural weather change (increasing of the world's temperature) and deteriorate the health of living creatures. One can only imagine the interconnections between the different threats to sustainability, being the cause and effects of each other.

The origin of sustainability dates back to the Brundtland report (1987) (Kuhlman and Farrington 2010). The main concern highlighted in this report was regarding the arising tensions between the human desires of a superior lifestyle and the restrictions put by nature. Thus, over the period, the idea of sustainability has been re-deciphered as enveloping three aspects specifically—social, economic, and environmental. Each of these three aspects needs to be examined on different but suitable parameters by professionals of different fields depending upon their areas of direct impact as the basic grounds for such a classification of sustainability.

While socio-economic sustainability focuses on several factors of social equality, citizen welfare, economic growth, and expenditures, the goal of environmental sustainability is to moderate the consumption of natural non-renewable energy sources and to create alternate sources of energy while diminishing contamination and damage to the earth. Thus, the intention behind achieving environmental sustainability is to address the issues of the present without compromising the limits of individuals to address their issues in the future. In light of this affair, it is reasonable to state that Renewable Energy Sources (RES) are viable direct arrangements for sustainable development as they are clean and non-exhaustible sources of energy. The next major concern regarding environmental sustainability is the lack of adequate technologies to capture and effectively utilize energy from renewable sources due to knowledge gaps. Such an issue is quite common in Less Economically Developed Countries (LEDCs) as in the case of Nigeria. Even though Nigeria has a huge potential for the growth of solar power due to high values of annual incident solar radiations, solar radiations are being wasted due to the lack of local investments in solar modules, PV arrays, and well configured infrastructural architecture. It is reported that solar modules covering 1% of Nigeria's landmass can generate up to 627,260 MW of electricity yearly as compared to the current generation of 6803 MW (Ndukwu et al. 2020).

Thus, nations need to focus on overcoming the connected web of challenges related to all three aspects of sustainability. In this regard, several data-driven emerging trends such as Artificial Intelligence (AI), Data Sciences, Internet of Things (IoT) integrated have revolutionized several industries and organizations and have a much larger potential in the future. Thus, data is seen as a resource powering up the informational economy.

From the relevance of this chapter, we will mainly be focusing on research concerning the applications of big data-driven technologies in achieving sustainable development goals.

But what exactly is big data? Big data as the name suggests refers to huge collections of data, i.e. can be anything varying from records, facts, statistics, audios, videos, photos, figures, etc. Having said that, big data is worthless on its own, unless we associate it with added values. We can explain this dubious property of big data with an analogy similar to that of currency notes. In their true sense, banknotes and coins do not have any physical significance than just being pieces of paper and metal chunks. However, humans have added value to the banknotes by authorizing them from legal authorities such as banks. Similarly, big data sets need to be processed and then examined for deeper insights relative to the field of impact.

What is Big data analytics? Big data analytics is effectively breaking down the data collections to information, using the information to increase the existing knowledge and finally attaining wisdom from knowledge (Tien 2013) to make more intelligent business moves for an efficient future. Hence, the idea of Big Data analytics is viewed as a development of Business Intelligence (Big Data vs. Business Intelligence vs. Data Mining|The Differences 2021), since it incorporates new sorts of data that were already too complex to even think about handling them with usual procedures (Sanila et al. 2017). Cost assumes a major role in demonstrating the significance of big data where organizations tend to achieve ideal costs for products to expand their benefit as well as to ease the cost-related concerns of the customers.

The motivation behind this chapter is to review and highlight the role of big data in accomplishing sustainable development goals through various sectors of a smart city. A smart city is essentially expected to integrate all important sectors of human interference having an impact on sustainability. Since big data is one of the emerging technologies since the past few years, this chapter also contributes to the existing research by reviewing previous research papers that integrate various other emerging technologies such as Machine Learning (ML)/Deep Learning, Internet of Things (IoT), etc. with big data analytics to present a holistic view of the current research in sustainable development. We will also emphasize on how professionals can leverage the data collected, recorded, and generated on large scales using big data techniques to move closer in achieving sustainable development goals. Close attention is also paid to the prevailing challenges and their possible solutions.

The rest of the chapter is divided as follows—Sect. 2 provides essential background about sustainability, renewable energy, big data analytics, and E-mobility. Following up with the background, a comprehensive review of the existing literature is done for the same in Sect. 3. Section 4, 5 and 6 summarize the literature review, discuss the limitations of existing literature, and explore the scope of big data analytics in sustainable development and e-mobility, respectively. Section 7 presents a case study based on electric vehicle (EV) charging data analytics performed for the city of Mumbai, India to identify issues in the existing charging infrastructure and prepare for its future challenges. Finally, Sect. 8 summarizes the objectives of this chapter.

2 Background

2.1 Scenario of Global Sustainability

Global Warming has been one of the most severe consequences of air pollution due to green-house gases. Recently, one of the most environment triggering news that surfaced across the globe indeed concerned the topic of Global Warming. Scientists from the UK have concluded that the Earth has lost about 28 trillion tons of ice in 30 years as a result of global warming (Earth has lost 28 trillion tonnes of ice in less than 30 years 2021). This comes after similar harsh remainders about severe glacier melting in Greenland were put up in several reports in the first half of 2020 (Snowden 2021). A relatively new threat to sustainability is the increased number of wildfires in several parts of the world including California in the USA, Uttarakhand in India and Siberian regions. The European Union hence estimated that the Arctic wildfires alone contributed to about 35% more carbon dioxide generation up till 24 August 2020, than the whole year of 2019 (Siberia heatwave: why the Arctic is warming so much faster than the rest of the world 2021). Increasing records of melting of glaciers and wildfires suggest to us that the temperature of the earth is ever increasing and hence a major threat to global sustainability.

Likewise, COVID-19 times have worsened the situation and made it even more challenging to achieve the 2030 agenda for sustainable development put up by the United Nations. Some positive impacts have also been noticed, such as reduced planned travels (which resulted in less vehicular GHG emissions and hence smaller steps towards better air quality), fewer traffic congestions (which resulted in less travel time for emergencies), increased hygiene levels in several households (which resulted in the prevention of various diseases, not just COVID-19), etc. However, negative impacts have affected the economy and the environment much more severely. More plastic and related wastes are generated due to the huge demand for PPE (personal protective equipment) kits, gloves, and masks. Usually, these materials are also discarded after one-time use and hence the amount of waste generated increased considerably. As per the Chinese Ministry of Ecology and Environment, medical centres in Wuhan delivered an excess of 240 tons of waste every day at the height of the pandemic, in contrast with the 40 tons of waste during normal occasions (The plastic pandemic is only getting worse during COVID-19 2021). Six months since the outbreak now, and the world already has 105 million COVID-19 cases with 25 million of them being active as on 5 February 2021 (COVID Live Update: 171,029,617 Cases and 3,556,677 Deaths from the Coronavirus—Worldometer 2021). With such figures, it is expected that the per day waste generation has crossed records by now. Alongside, a great recession is being forecast, bringing the world's economies down to their knees, especially the Less Economically Developed Countries (LEDCs). The loss of jobs and the temporary shut-down of various offices has resulted in the reduction of annual GDP of several developing countries including India. The World Bank had already estimated about 3.1% loss in the Indian GDP for the fiscal year 2020–2021 (Data|An estimated 12.2 crore Indians lost their jobs during the coronavirus lockdown in April: CMIE 2021). Making the situation worse, the official data released by the National Statistics Office (NSO), India on August 31, 2020, revealed the first economic contraction of India in 4 decades, as much as 23.9% decrease in the April-June quarter of 2020 as compared to the same quarter of 2019 (Team 2021).

Figure 1 represents the Global Sustainable Competitiveness Index (GSCI) for the year of 2019 which measures the seriousness of nations in terms of sustainability in a coordinated way. It is determined dependent on 116 quantifiable, quantitative markers acquired from many reliable sources, for example, the World Bank, the IMF, and diverse UN organizations (The Global Sustainable Competitiveness Index 2021). Sweden tops the list with a score of 60.6. Finland and Iceland bag the 2nd and 3rd positions with scores of 59.5 and 57.3. The United Kingdom is on the 17th rank and one of the countries to cross the 50 mark with a score of 52.8. USA (34), China (37), & Nepal (64) follow up with scores of 49.1, 48.5, and 45.6, respectively. India comes on the 130th rank with a score of 39.5. India's neighbour Pakistan secures the 145th positions with a score of 38.3 whereas the last country in the rank-list is the Bahamas with a score of 30.5. Figures suggest, there is a long way to go!

Thus, due to the largely disruptive impact of COVID-19, the world is in an alarming position. More judicious utilization of resources is necessary to bring back the normal. Developing countries especially need to manage finances and resources

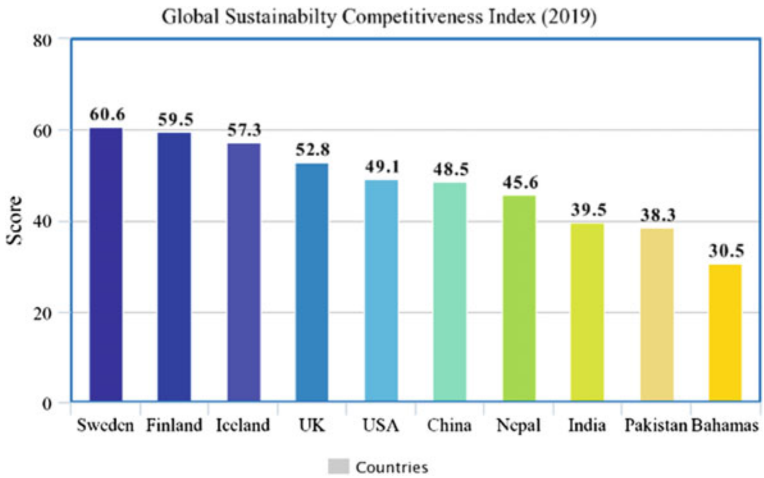


Fig. 1 The GSCI indexing scores for major countries of the world in the year 2019

whereas the developed countries need to severe down the waste generation and GHGs emissions.

The threats to sustainability are larger for India. India is the world's second-largest country by population and among the most densely populated countries. The per capita income of India remains among the lowest in the world even though India is the world's largest economy. Thus, the results of 'un-sustainability' not only affect the demands of the future, but also the public health of the present. It also puts enormous pressure on carrying capacity of the natural resource system as well as the inability of the environment to absorb the waste generated.

As a conscious aspirant, India has played a lead role in tackling climate change and putting forward policies that align with sustainable development goals of 2015. That being the case, endeavours have been made to include all partners—from people in general and private divisions—in the framing process of efficient policies, strategies to be implemented, and helpful practices. The international report on sustainable living—'Greendex' measures how consumers all over the world are reacting to sustainability concerns (Greendex Ranks 18 Countries Based On Sustainability 2021). Indian consumers occupy the second spot on this index, particularly receiving high grades in the sectors of housing, transport, and food choices (INDIAN CONSUMERS: Second place, n.d.).

2.2 *Green Renewable Energy*

Around 75% of electric generation systems in India use fossil fuels (Growth of Electricity Sector in India 1947–2020 2020) (coal, diesel and natural gas) as their major source and almost all vehicles run on crude oil derivatives such as petrol and diesel or on fossil gas (Fig. 2 further represents the shares of other sources of electricity generation in India). These non-renewable exhaustible sources of energy are estimated to last for not more than a few more decades (if the rate of consumption doesn't lower with respect to the rate of natural production). Thus, due to the constantly increasing need for energy in the last decade and climate change issues, RES implementation is vital. As of September 2020, India has a total installed grid-interactive renewable power capacity of 89.23 GW (web.archive.org 2018). Currently, out of a total of 89.23 GW of renewable energy 38.12 GW from wind energy, 36 GW from solar, 10.1 GW from bio-power, and 4.73 GW from small hydro plants.

Cities are found to be responsible for about 70% of energy-related CO2 emissions (Science 2018). Since renewable energy generation has gained much popularity in the past decade, several cities across the world have switched to these clean sources of energy, partially and in some cases entirely. With cities such as Brasíla, Burlington, Sydney switching to 100% renewable energy (www.power-technology.com, n.d.), it would be appropriate to claim that renewable energy can be used as a major power source in several leading cities and there is an immense potential in cities to lead on building a sustainable economy. India is taking baby steps in the same direction, Jaipur and Bengaluru being the first Indian cities in world's top 10 renewable smart

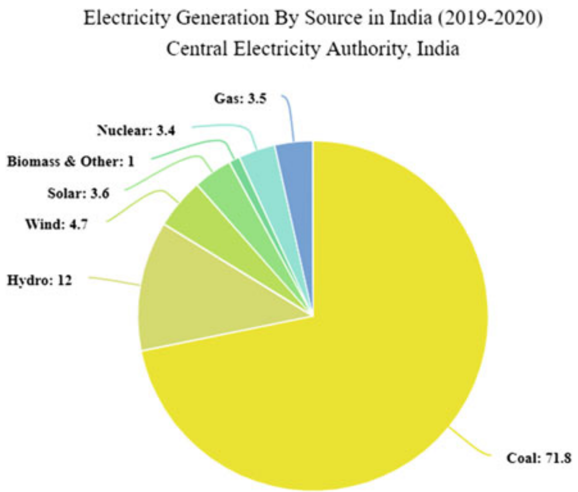


Fig. 2 Sources of electricity generation in India with their respective shares in 2019–2020

cities list bag the fourth and the sixth positions, respectively (www.power-technology.com, n.d.). India is also expected to be the largest contributor to the renewables by 2021, with the country's annual additions doubling from 2020 (EDF Energy 2015). Recently, Asia's largest solar power plant was inaugurated by Prime Minister in Madhya Pradesh's Rewa with a capacity of 750 megawatts. With the deployment of this power plant, it is estimated that the reduction in carbon dioxide emissions will reduce by 15 lakh tonnes annually (The Indian Express 2020).

2.3 Big Data

Data is being generated at a fast rate; almost more than 5 Exabyte per day increase due to use of emails, audio, video and data streams, health information, various queries, social networks, scientific data, and on-going mobile phone applications data (Sanila et al. 2017). It involves highly delicate multiple datasets owing to different aspects such as personal data, organizational data, environmental data, data of transport systems, and economic data. Thus, the applications of big data analytics in several aspects of day-to-day life as well as in several industrial, organizational, and business aspects is where humans benefit from the big data.

Large data production in the form of government records and surveys can be termed as universal data sets since they provide information related to both—the city and its citizens on a whole. However, currently one of the most popular and common data sources is the internet. Internet websites are not only based on data but also record data in the form of user activities (Einav and Levin 2014) and hence largely contribute to big data sets. Almost, 59% of the world's population is actively using the internet on an everyday basis (Johnson 2021). From personal data such as search history, text messages, emails, photos to relatively public data such as online shopping, registrations, social media, interactive websites, etc. each activity is being recorded in real time for every single user. Imagine, the amount of data generated by a single person in a day, and then imagine how much the whole world is generating daily! This is the kind of big data we are talking about. Other specific sources of data vary with the areas of concern because internet activity records majorly qualify as social and economic data, not environmental data. While the most common sources of data addressing environmental concerns include photos, SNS (social networking sites) data, survey records, and machine data (which is then put up on the internet websites), the continuous technological improvements in distributed technologies such as the IoT, GIS (Geographic Information System), inductive loops, smart meters (AMI) and particularly in intelligent electronic devices (IEDs), sensors, and actuators, have led to the production of huge amounts of data (Junaidi and Shaaban 2018). Similarly, economic data sources have evolved. With the introduction of scanners, scanner data seems to have revolutionized sales and marketing by providing detailed information such as quantity, type, prices, store ID, etc. about the goods and products sold. Scanner data and electronic inventory records are the current major sources of economic sales data (Einav and Levin 2014). Big data applications have a large role in improving environmental sustainability in the realm of smart cities and their activities. They have the potential to serve a variety of domains of smart cities such as power grids and/or smart grids, transport, traffic, urban design and planning, infrastructure monitoring and management, etc.

Although, data storage is one of the major challenges faced in big data analytics, relatively big organizations have achieved convenient data storage facilities for current as well as future use through constructive planning. For instance, through the GitHub Archive programme, several terabytes of data comprising of mainly open-source software were safely stored for use by future generations in a vault in the Arctic for as much as 1000 years (Months and Weeks 2020). Also, the demand for stream processing is increasing since most of the data generated in the modern world is real time (Vanani and Majidian 2019). When big datasets are combined, cases might be distinguished by the integration of data which then provides countless opportunities to apply reasoning in order to think so as to grow our natural organizational structure (Wu and Chen 2017).

Big data characteristics allow the utilization of big data methods to extract value from the business. The 5 V's are Volume, Velocity, Variety, Veracity, and Value (Amaro and Pina 2017). Figure 3 represents the basic underlying knowledge behind the 5 V's of big data.

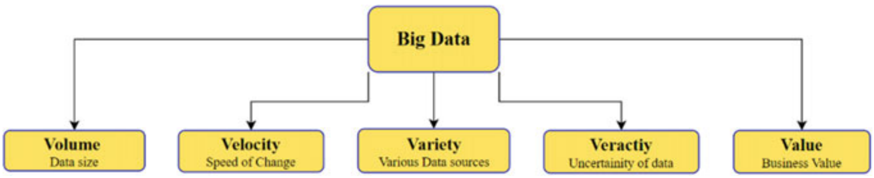


Fig. 3 The 5 V's of big data with basic underlying knowledge

Usually, when dealing with a limited amount of data with less variability as compared to big data, the MySQL database is used. It is an open-source relational database management system (RDBMS) that collects, cleans, analyses, and visualizes the data using Structured Query language (SQL). However, for big data sets possessing the 5 V's, the operating performance for transferring the data to RDBMS is greatly compromised. The current industry standards for decision support system are based on Data warehousing (Nitesh Jain 2015). Data warehousing is simply leveraging the big data produced in repositories, websites, weblogs, market trends, transition logs of customers, etc. by industrial data analysts. Similar to the drawbacks faced by the traditional RDBMS in querying through big data, the algorithms of data warehousing supported only small data sets where increasing the sample size of data sets would take a high turnaround time.

Hence, an open-source framework capable of handling big data sets at quicker throughputs were needed. This led to the evolution of several big data handling frameworks, databases, data warehouses, etc. such as Apache Hadoop, Apache Cassandra, Apache Spark, Google Big query, MongoDB, Xplenty, and a few more. Table 1 derived from (Apache.org 2019; Softwaretestinghelp.com 2019) gives a basic overview of a few commonly used big data tools and technologies and the languages supported by them.

Hadoop is known as one of the first and the most efficient big data software frameworks to date. The basic idea behind Hadoop and several others is using a cluster of computers instead of increasing the computation power of one computer. Hadoop is based on the idea of distributed computing where distributed frameworks use a cluster of computers for enormous amounts of constantly increasing data. Bill Gates thus popularized the idea of business at the speed of thought enabled by Hadoop. (Nitesh Jain 2015).

The basic two advantages of clustering are:

- Clusters are commodity hardware, cheaper than one expensive server (studies have shown that Hadoop clusters can give 10 times more throughput at 1/10th of the cost in certain situations).
- The license fee of traditional RDBMS is expensive, but Hadoop's software license is free and open source.

Hadoop core provides a layer between the user and cluster of machines under it. It comprises of three core components. The Hadoop distributed file system (HDFS), the programming framework or Map-reduce and Yet Another Resource Negotiator (YARN). Through Map-reduce, the code is moved to the data, unlike of what happened previously, where data was sent to the code. YARN assigns computational resources for execution of an application (Datadog 2016).

What happens inside Hadoop is an interesting question to answer. The programming or the MapReduce framework enables users to have full control on the processing of input datasets (Dittrich and Quiané-Ruiz 2012). A file when given as input, is broken down into smaller portions known as blocks. The blocks are then replicated and distributed over a cluster of computers. It is Hadoop that manages the

Table 1 Big data tools and technologies

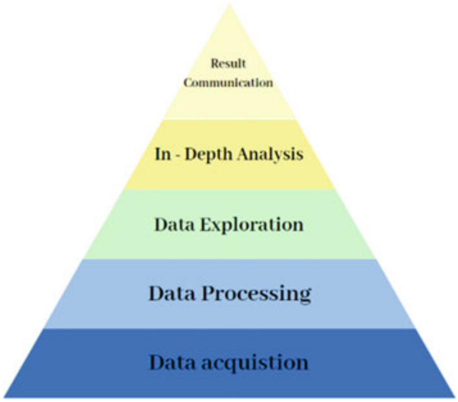
Name	Owner	Type and description	Language
Hadoop	Apache Software Foundation	Java-based framework which runs applications on clusters of commodity hardware • Basic Components—Hadoop Distributed File System (HDFS), YARN, Hadoop MapReduce & Hadoop Common • Additional Software packages from Apache capable of running alongside Hadoop- Hive, Pig, HBase, Spark etc	Written in Java. Java is also preferred for programming however other languages such as Python, Scala, and SQL are supported depending on the type of additional software packages
Cassandra	Apache Software Foundation	NoSQL database management system which handles data across many commodity hardware	Written in Java, employs Cassandra Query Language (CQL)
Kafka	Apache Software Foundation	Stream processing software platform	Supported languages—Scala and Java
Big Query	Google	Serverless data warehouse for scalable analysis	Supports ANSI SQL
MongoDB	MongoDB Inc	Document oriented database program	Written in C++, Go, JavaScript, Python. Supports all popular programming languages
Xplenty	Xenon partners	Scalable cloud platform to integrate and process data from various sources	Supports SQL, Python and R

distribution, division, and management of the blocks, hence easing the task for users. Overall, programmers just need to write the Map and Reduce codes and the rest of the work is done by Hadoop distributed file system (Udemy, n.d.).

2.4 *The Analytics Procedure*

By now, it is fair to comprehend that the basis of any organization is formed by data analytics and all big data applications deal with a similar procedure of data acquisition, cleaning, processing, exploration, in-depth analysis and finally communicating the results for smart decision-making. In this section, we attempt to give the reader a general overview of the stages of a data processing procedure which then forms the foundation for implementation of various ML algorithms.

Fig. 4 The pyramid represents steps involved in a data analytics procedure in an orderly manner



The data acquisition process involves the collection of raw data from all possible data sources and acquires the data in usable formats such as.csv,.json,.xml, etc. The stage of data processing usually needs the analyst to examine and understand each column. Data processing involves two stages where the first stage also known as data pre-processing aims to prepare data for analysing by defining a common standard (Amaro and Pina 2017) or by using the concept of 'virtual objects' (Linder et al. 2017) to solve compatibility issues. Further, in the second and final stage of processing, it is important to filter out/replace errors and missing values to refine the database. The data exploration stage incorporates splitting, grouping, and plotting the data in different forms of visualization to identify patterns, trends, and correlations. At this stage, the data is ready for an in-depth analysis using predictive models based on big data techniques such as ML/Deep learning for several purposes such as clustering, pattern recognition, regression, classification, feature selection, or time series forecasting, etc. Finally, from visual observation of graphs, charts, and figures it is also obvious how communication might be seen as the most important step in the analytics procedure to gain crucial insights regarding the inputs, outputs, profits, efficiency, and productivity. Unless and until the analysed data is not communicated properly to the concerned professionals or the common man in several cases, data seems to have no associated value. Figure 4 gives a visual overview of the steps performed in the data analytics procedure.

2.5 E—Mobility

Currently, the energy sector in transportation is majorly oil driven. Oil being a fossil fuel is a non-renewable source of energy hence limited in nature. In India, diesel and motor gasoline represented 90% of the final energy consumed in the transport sector, while jet kerosene represented 8% and electricity 2% in 2004. Diesel is the most used form of energy, with a share of 66%, and motor gasoline representing 24% (De la Rue du Can et al. 2009).

Apart from limited availability concerns, a major concern is the air pollution caused by the emission of harmful pollutants such as GHGs, VOCs (Volatile Organic Compounds), NOx, CO, etc. by combustion of oil in vehicles, which is also harmful to the health of living beings. In the United States, cars and trucks collectively emit around 24 pounds of carbon dioxide and other global warming gases for every gallon of gas (Union of Concerned Scientists 2014). About five pounds comes from the extraction, production, and delivery of the fuel, while the great bulk of heat-trapping emissions—more than 19 pounds per gallon—comes right out of a car's tailpipe (Union of Concerned Scientists 2014). Hence, the transportation advancement sector is now driven by the need for cleaner, RES such as electricity. Subsequently, transportation electrification is generally viewed as an appealing solution for reducing the reliance on oil for transportation and environmental impact of road transportation (Fig. 5).

The first step towards transportation electrification was the invention of electric vehicles (EVs). The number of PHEVs sold as on December 2020 in major global markets were about 13 million with China having the maximum sales every coming year, followed by Europe, USA, and other major countries (Wikipedia Contributors 2019a). India experienced sales of about 1.56 lakh units which is comparatively lower than sales in other leading countries but still an increase of about 20% from last year's sales (BloombergQuint, n.d.).

The numbers are huge, however not even close to taking over the ICE (Internal Combustion Engine) vehicle sales in the world. It is estimated that the world has surpassed the 1.5 billion mark in the number of vehicles on road (Chesterton 2018), which is quite a lot more than the total number of electric vehicles in the world.

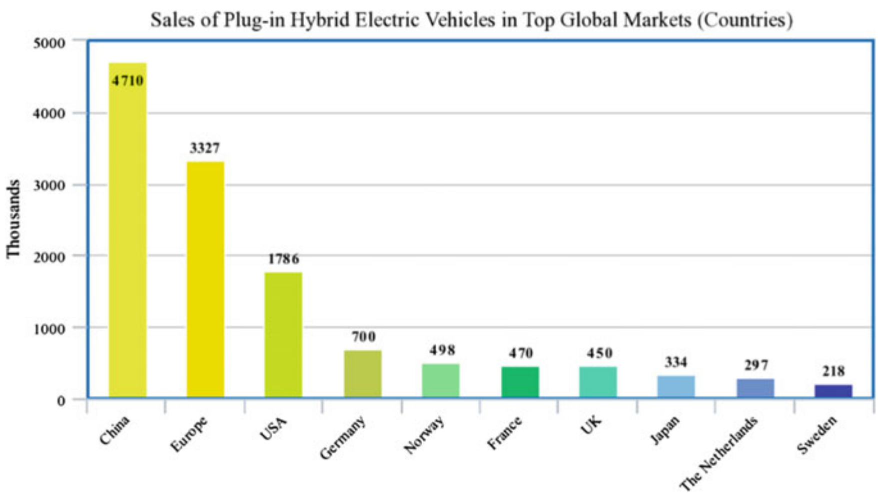


Fig. 5 Number of PHEVs sold till the year 2020 in major countries of the world

Thus, suggesting to us—‘*the increasing but not yet enough*’ EV sales in the world in order to achieve environmental sustainability. This largely directs us to the EV ecosystem which consists of several factors that determine the extent of adoption of EVs by users. Building a profitable EV ecosystem in a city implies building a network of service providers and businesses (Guardknox, n.d.) in the EV domain. An EV ecosystem is classified into four major service provider and business owner components based on their area of application (Fig. 6). Further, each component can be classified based on attributes of concern. The EV ecosystem is still on growth in the world and several current technologies and practices need to be on the radar of the government so as to popularize EVs among the users and make them think twice before they consider purchasing an ICE vehicle.

When an EV is manufactured, the technical component of the EV ecosystem plays an important role in determining the efficiency and benefits gained through an EV as a mode of transport as compared to oil-based modes of transport. Currently, lithium-ion batteries are the most popular type of EV batteries due to their higher specific energy (López-Ibarra et al. 2020) relative to the other battery types. However, it is found that batteries have much shorter lifetimes than electronic power systems with about 5–8 years of warranty (López-Ibarra et al. 2020) and (EDF Energy 2015). Shorter lifetimes imply investments in new batteries as much as after every 8–10 years of use. New battery investment costs along with the original total costs of ownership of the EVs (where batteries account up to around a quarter of the total EV cost López-Ibarra et al. 2020) are serious concerns for EV users worldwide. EV batteries also play a major role behind ‘range anxieties’ and ‘charging time anxieties’ of users. Such user anxieties can be solved by improving several battery features such as capacity, State of Charge (SOC), the integrated battery management system, and lifetime. However, manufacturers need to keep in mind another set of battery features while production. The batteries shall maintain an affordable cost range, should not be bulky or heavy as compared to the EV and finally have lesser harmful environmental impacts while disposal. The technical component also includes manufacturing of batteries with safer operations, in order to avoid risks.

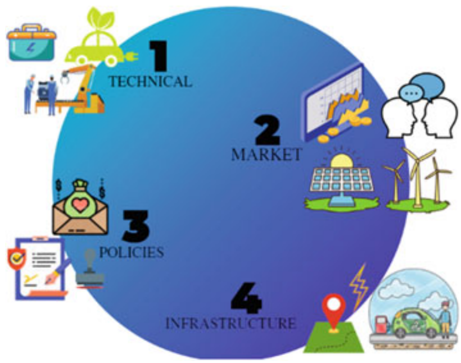


Fig. 6 Four major components of an EV ecosystem

Once, the EV is manufactured, it is time for its widespread acceptance by users through market trends and policies (Vikaspedia.in 2020). Market values along with policies play an economical role in popularizing EVs. The EV market is expected to assist EV buyers by answering queries related to consumer perceptions and emphasizing on long-term benefits through testing and certification, servicing and affordable overall costs. The Market is not just limited to factors directly related to an EV, but also include factors like quality of electricity in the region and market for storage of electricity. Several policies by the government regarding taxes, subsidies on sources of electricity, or electricity prices on a whole are expected to help lift the economic anxieties of citizens into purchasing EVs.

Finally, the last component of an EV ecosystem playing a crucial role in widespread EV acceptance is the EV charging infrastructure (Vikaspedia.in 2020). The charging infrastructure deals with appropriate locations of charging stations, allotment of charging points to charging stations and an adequate number of charging points. EV charging infrastructures are much more sensitive to the placement of chargers and charging stations as compared to fuel refilling stations for ICE vehicles due to one major issue—higher charging times. On average, even the fastest chargers require about at least 30 min to charge an EV and the charging times vary with the type of EV. Subsequently, higher charging times also require proper placement of charging stations for an EV user to be able to utilize his time while the EV is charging. Further, to assist issues related to high charging times, charging station service providers need to successfully implement the new-age technologies such as battery swapping and battery replacement. However, such practices require the batteries and EVs to undergo another set of standardization procedures.

Electrical Vehicle Supply Equipment (EVSE) commonly known as an EV charger, is not just a charger. An EVSE control aims to solve compatibility issues between EVs and different types of chargers. EVSE basically enables any EV user to charge from any outlet. It is equivalent to a settlement between the EV and the charger to keep the EV safe while charging using two-way communication between charger and EV (Green Car Reports, n.d.). At the same time, EV charging can be classified by the 'types of charging' based on the nature of contact/communication between EV and the charger. Thus, the two types of charging infrastructures based on the type of charging are—Conductive charging infrastructures and Inductive Charging infrastructures (Khalid et al. 2019). Since conductive charging involves direct metal to metal physical contact between the EV and the charger, it is more efficient & robust as compared to inductive charging (based on the principle of mutual induction). Hence, conductive charging is much more widely accepted and can be broadly classified based on the type of current, charging times, and voltage and current levels in Table 2 derived from (Khalid et al. 2019; Sharma et al. 2020).

Another major limitation in the charging infrastructure is the quality of electricity provided by the grid to the vehicles keep fluctuating as discussed earlier. These limitations thus tend to spark trust issues in the user about the time taken for EV charging and the cost generated. A grid management system might help in reviewing over-power flows in the grid so as to even out the fluctuations and avoid surges.

Table 2 Types of conductive EV chargers with specifications

Charger name	Type of current	Time taken to charge	Voltage and current ratings
Level—1 Charger	AC Slow	8–16 h for full charge	120 V, 15–20 Amp
Level—2 Charger	AC Slow	4–8 h for full charge	240–400 V, 40–80 Amp
Level—3 Charger	DC Fast with CHAdeMo technology	10–15 min for 80% charge, remaining 20% need constant voltage which consumes time	208–600 V, 200 Amp

Likewise, EV management can be given a more intelligent approach by several practices such as carpooling, ridesharing, self-driven cars, etc. A few studies insist on the large role of the influence of other consumer's perception in local EV adoption, i.e. 'word of mouth' effects (Cai et al. 2013) which simply encourages a non EV user to think about personal EV adoption. EV adoption in different regions or countries is also dependent on the average annual income of citizens of that country. For instance, most Americans and Chinese citizens can afford EVs and hence EV culture is soon to be predominant in those countries. However, EVs are still a foreign concept in developing countries like India. Having said that, income alone cannot be held responsible for EV adoption among citizens. For instance, the average annual income of citizens of the UK is among the highest in the world, but EV adoption is not exactly as predominant in the UK as in China. Hence, a customer service driven approach also plays an important role in transforming the EV market. It is basic to comprehend that customers resist change whenever introduced with a new technology or innovation. Thus, it is important that transportation electrification programmes offer genuine and tangible advantages to the EV drivers and afterwards plainly convey those advantages. Utilities can furnish both private and business clients with projects and administrations that energize and reward the adoption of EVs in their administration regions by leveraging best activities of the past and keeping in mind the client's point of view all through the advancement procedure (www.tdworld.com, n.d.).

3 Literature Review

It is crucial to study literature samples regarding recent advancements utilizing big data in several components of a smart city to not only analyse the extent up to which the big data is useful to us but to also raise relevant questions for future research directions.

3.1 Smart Cities

We begin our study by a general understanding of what urban areas under the umbrella idea of smart cities are expected to offer to its citizens.

Urban areas have held on to the Information and Communication Technologies (ICTs) as an advancement technique, hence bringing about the concept of a sustainable smart city. Through specific components of a smart city (Fig. 7), a big-data-driven smart city is expected to not only bring us closer to achieving the goals of sustainability and easing tasks for residents of the city, but also aims to address concerns related to socio-economic sustainability such as disaster detection and management, event detection, resources management, health monitoring, safety and security, profit maximization, etc.

A smart city can take several different definitions depending on the area of focus as classified in Fig. 7. For instance, the term smart city has been differently categorized in several pieces of literature such as 'intelligent cities' (Kominos 2002), 'wired cities' (Dutton et al. 1987), 'sentient cities' (Shepard 2011), or 'digital cities' (Gerhard Goos et al. 2000). (Kitchin 2013) studied the relationships of these terms with urbanization and concluded that these terms are utilized with a specific goal in mind to conceptualize the connection among ICT and contemporary urbanism, yet they mainly focus on the impacts of ICT on urban structure, procedures, and methods of living. (Gerhard Goos et al. 2000) further extensively classifies smart cities into two distinct yet related understandings with regards to what makes a city 'smart'. From one perspective, the idea of a smart city presented by (Khan et al. 2017) is—A city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, seaports, communications, water,

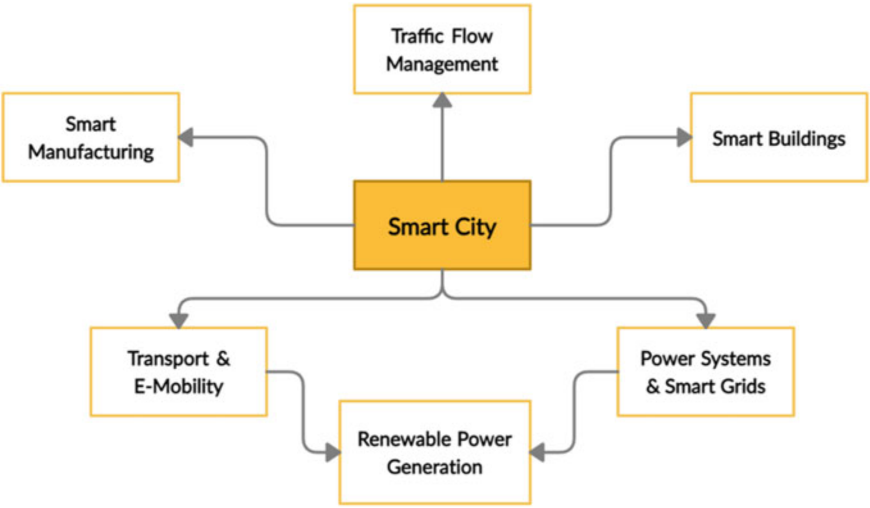


Fig. 7 Classification diagram of a smart city representing components of a smart city

power, even major buildings, can better optimize its resources, plan its preventive maintenance activities, and monitor security aspects while maximizing services to its citizens through data collected through smartphones, computers, GPS (Global Positioning System), sensors, cameras, and even people. But from the other perspective (Khan et al. 2017) has well defined the components of a smart city, such as mobility, governance, environment, and people as well as its applications and services such as healthcare, transportation, smart education, and energy systems.

The concept of IoT introduces the integration, monitoring and data collection processes (through sensors, actuators, smartphones, etc.) of several related objects working together for a specific task. While the IoT environment is vital for the development of a smart city, it is also responsible for the generation of huge amounts of real-time data in raw form which may be structured or unstructured, i.e. big data. (Rathore et al. 2018) in their study aim to collect real-time data concerning several domains of a smart city such as smart home, smart parking, smart vehicular traffic, weather and water system, surveillance and safety, and environment to harness and link the big data sets. Analysing and Visualization of these data sets is further expected to influence the decision-making processes in urban planning and hence in the development of a smart digital city.

Energy Management plays a very important role in the context of smart cities to substantially optimize the processes related to energy—generation, transmission, distribution, and utilization. It is of crucial importance to manage activities related to energy production and consumption not only to efficiently generate energy but also to leverage the generated energy. In the coming sections, we will emphasize on each of these aspects of energy management with the introduction of data sources and utilization of data-driven technologies.

3.2 Power Systems and Smart Grids

In this section, we will be emphasizing on 'Power Systems' that deal with the generation, distribution, and transmission of electricity through several different components. However, current power systems face a few distinct challenges associated with economic viewpoints (efficiency, productivity, affordability, system stability, and reliability) as well as in the successful implementation of big data analytics procedures (such as effective data acquisition and storage, data curation methodologies, how to exactly use the stored data to extract business value and how to reduce privacy issues). Eventually, once these challenges are dealt with big data analytics can again be utilized for the improvement in several operational sectors of power systems. Real-time optimization of power system network, precise forecast of load demand, consumption patterns analysis making way for new services, and pricing strategies are some of the areas that need assistance and big data seems to make its way (Zhou et al. 2016). Thus, the smart grid was conceptualized considering it to be a promising solution both from the consumer point of view and the industry point of view (Amaro and Pina 2017).

The introduction of smart grids employed advanced digital information and communication technologies which enabled the usage of information technologies like data analytics, forecasting, classification and clustering algorithms (Zhou et al. 2016) on the large amounts of smart grid big data. Apart from benefiting the industry owners, a smart grid also helps in rapid detection & restoration of faults and reducing the cost for consumers (Moradi et al. 2019). Smart grids are thus considered a promising solution both from the consumer point of view and the industry point of view.

A few intelligent devices currently being used in smart grids for efficient data collection and processing throughout the whole process of power flow include

- Phasor Measurement Units (PMUs) which have now replaced SCADA Systems due to higher sampling rates of PMUs (Junaidi and Shaaban 2018).
- BAS (Building Automation System)
- Sensors and Thermostats.

Similarly, major sources of data described in (Amaro and Pina 2017) are

- AMI data (smart metres)
- Distribution automation data (grid equipment)
- Third-party data (off-grid data sets)
- Asset management data (firmware for all smart devices and associated operating systems).

  Other sources also consist of
- GIS big data and weather big data (Voivontas et al. 1998; Jakubiec and Reinhart 2012).
- IED (Intelligent Electronic Devices): It integrates at least one microprocessor that can receive or send data/control from or to an outer source (Junaidi and Shaaban 2018).
- Electricity market data (Liang et al. 2018).

Further, the energy management system for smart grids is found to be more complex as compared to power systems. Ancillotti et al. (2013) in their study, described a smart energy management system as the new technologies and models that are fundamental for the well-off deployment of a smart grid and can be classified into three components

- Real-time wide-area situational awareness (WASA: a network of PMUs) of grid status through advanced metering and monitoring systems.
- Consumer's participation through home EMSs, demand response (DR) algorithms, and vehicle-to-grid (V2G) technology.
- Supervisory control through computer-based systems.

3.3 Smart Buildings

Out of the many energy-consuming aspects of smart cities, smart buildings turn out to be one of the biggest ones. Not only energy consumption rates but emission rates of GHG gases are also considerably high from smart buildings. Statistical studies have been performed and observed regarding energy consumed in the building sector. The building sector is one of the largest energy consumers, accounting for approximately 20–40% of the global energy usage (Chou et al. 2016; Linder et al. 2017) while generating 30% of all CO_2 emissions (Ancillotti et al. 2013); energy consumption by this sector is continuously increasing because of urbanization, rapid economic growth, rising income, and growing population. EIA's *International Energy Outlook 2017* (IEO 2017) projects that by 2040, the fastest growth in buildings energy consumption will occur in India as compared to all other countries. An average increase of about 2.7% per year delivered energy consumption for residential and commercial buildings in India is expected from 2015 to 2040 which is more than twice the global average increase. Figure 8 represents the country-wise annual percentage growth in the building energy sector (www.eia.gov, n.d.) (Fig. 9).

Major studies till date have concentrated on setting up a framework for anticipating the pattern of future power utilization through stochastic strategies and regression analysis to encourage effective usage of electricity by end-users (Lee et al. 2017). In this section, we will review some of the frameworks based on big data-driven technologies that have been described in the existing literature utilizing the context of an intelligent building. An intelligent system in the context of an intelligent building is composed of three levels: The infrastructure level of the input data, system infrastructure level and the level of services (Daissaoui et al. 2020). These three levels are going to form the basis of the majority of the big data models discussed.

A Smart Decision Support System (SDSS) proposed by (Chou et al. 2016) integrates smart grid big data analytics and cloud computing for building energy efficiency. Based on layered architecture the SDSS contains a data access layer followed

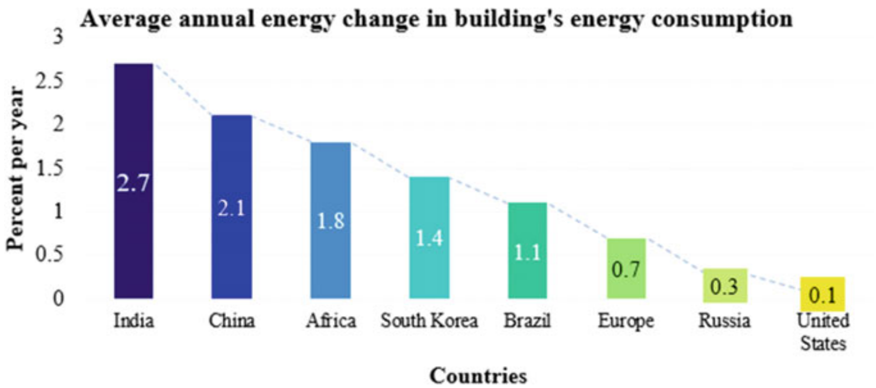


Fig. 8 Annual increase in building energy consumption for a few major countries/continents

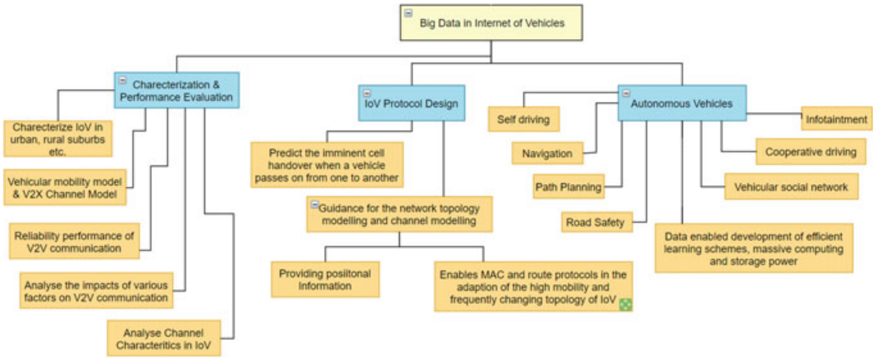


Fig. 9 Several domains of Big data analytics aided IoV

by an integrated analytics bench and a web-based portal. Advanced AI techniques and dynamic optimization algorithm are possibly coordinated behind the analytics bench to empower precise prediction and optimization of energy consumption. End users can minimize their costs by the best-case automatic operation of appliances by SDSS. The general flow of this framework includes four major steps

i. Collection of real-time data from AMI (smart metres).
ii. Wi-Fi and Bluetooth act as data communication platforms.
iii. AI & multi-objective algorithm for data analysing and
iv. Web-based tech visualizes the optimal operating schedule.

Additionally, (Linder et al. 2017) introduces Big Building Data (BBD) as a platform important for the paradigm shift from BMS to the web of buildings. BBD is ingestion, handling, and sharing framework ready to scale up to the Big Data desires for smart building situations. It provides data monitoring, anomaly detection and prediction, and renovation tools for users (building user, building owner, architects and Engineering consultancies). Another opportunity of this approach is BMS revolution by decoupling of sensors, data storage, and application levels supported by IoT and Web of things. Smart controls tend to address various sources of consumption such as cooling, heating, electric lighting, ventilation, solar shading, electric appliances, etc. We observe that such technologies rely more and more on the gathering of large amounts of data from multiple sensors, actuators, and dedicated networks.

Daissaoui et al. (2020) in their study, focus on the importance of IoT-based analytics. Nowadays, IoT-based big data frameworks are indispensable from smart buildings point of view. Two inseparable components from this point of view are IoT, which is made of all the associated sensors and the capacity condition for the data produced by these sensors. This framework is based on the recent research proposed to describe uniform meta-data for modelling buildings. These practices use sensor ontologies, subsystems, and connections, guaranteeing interoperable and compact applications.

3.4 Transport Systems

As of 2017, 1.43 million vehicles are registered in the world (Wikipedia Contributors 2019b) consuming a huge amount of crude oil derivatives (fossil fuels) and producing a considerable amount of air pollution which is also risky to human and animal health. While the solution relies on switching to alternative fuels such as biofuels, electrolytic hydrogen, and non-fossil fuel-based electricity (Wu and Chen 2017), there still exists a need to critically identify the current situation, evaluate the problems and to design appropriate strategies to achieve sustainable development goals. This is where big data analytics steps in.

The job is not just limited to reduction of emission but also to limit the usage of natural resources and to make the most out of the utilized resource. Wang et al. (2018) in his study elucidates into why and how big data and emerging technologies can help minimize both the harmful carbon emissions and the assets utilized in the transport sector. The report gives a general idea that the current gains in energy efficiency are much lesser than the efficiency gains observed in the Information Technology (IT) sector (as expressed in 'Moore's Law'). Thus, proposing that more utilization of IT in the transport sector will deliver enormous profits in sustainability.

Transport big data—One way to look at smart urban transportation management is to consider it as a multifaceted big data challenge which may incorporate tracking and management of every aspect of the vehicle body and engine/battery maintenance, warehousing, routing and mapping, communications, etc. by big data collection from all relevant sources such as road traffic index, metro operation data, road accident data, passenger card data, elevated off-ramp data, bus real-time data, microblogging traffic data, air quality status data, weather data, GPS data, FCD (Floating Car Data: to detect traffic flow speed) and LBS (location-based service) data (Miller and Harvey 2011; Wang et al. 2018).

Personal Travel Assistants (PTAs)—Currently, the most common platforms utilizing transport big data are PTAs. PTAs are expected to make use of transport big data to assist the user for the most optimal modes of transport and fuel-efficient routes based on basic user inputs. The PTA architecture generally consists of data ingestion, storage, analytics, and optimization modules to put forward a holistic as well as a sustainable approach for each trip (Miller and Harvey 2011; Wang et al. 2018).

Automated Vehicles (AVs)—The potential use of big data in AVs is particularly in object detection and deep learning approaches to make wiser and more accurate decisions.

Transport Technology and Mobility Assessment (TEMA)—A data processing platform expected to govern several road transport policies including the deployment of EVs, shift from oil to electrical energy, evaluating emissions, outflows, and so forth. To leverage the big data generated, a sample of vehicles was taken for analysis and it is found that half of the vehicles are parked for more than 90% of the time and 78% travel up to 50 km/day, 9% exceed 100 km/day and 3% exceed 150 km/day.

The figures also suggested a large potential for deploying battery electric vehicles (Gennaro et al. 2018).

Internet of Vehicles (IoV)—It is highly expected that very soon modern vehicles will have the ability to connect using radio technologies and exchange data with the surrounding environment under the dome of Internet of Vehicles (IoV) (Xu et al. 2018). Thus, this huge network of connected vehicles is not only consuming big data but also generating big data, hence explaining the reciprocal relationship between IoV and big data and how both assist each other in their respective domains. While IoV is solely assumed to be assisting in the generation of big data, it also supports in transmitting, storing, and computing big data. Similarly, (Xu et al. 2018) explored the role of big data in assisting IoV in several domains such as performance evaluation, characterization and communication protocol design utilizing measurement data, trace data, trajectory data, field data, vehicle movement data, senor data, traffic data, map data, etc.

A few models have also been developed in the recent past based on big data technologies to achieve sustainable environment goals in transportation. (Wu and Chen 2017) presents a model on hybrid data analytics to tackle critical issues of green transport systems as well as to generate valuable operational strategies for transport firms to achieve environmental sustainability. The model thus presented highlights three main stages

- Business analytics: Data collection and exploration, Data selection, Data pre-processing, Data mining, and Data transformation
- Hybrid data mining: Topic mining and association rules.
- Applies text mining of big data analytics.

3.5 *Electric Vehicles*

Electric Vehicles come with several significant benefits majorly—zero emissions and the capability to depend fully on RE sources. Despite such favourable outcomes of an EV, EVs are a rare sight in most countries of the world, especially India due to three major limitations of EVs. EVs are known to be much more expensive with shorter driving ranges and much greater charging times as compared to fuel vehicles. Thus, from an EV deployment point of view, it is necessary to review on the existing articles related to both—the technical aspect as well as the infrastructural aspect to tackle challenges associated with the mass adoption of EVs among users.

Technical Aspects—An EV battery plays the most important role in the technical aspect of an EV ecosystem. A Battery Management System (BMS) is an electronic system that controls a rechargeable EV battery (Yevgen Barsukov and Jinrong Qian 2013) by providing safety assurance to the battery, calculating and reporting data, monitoring the battery's state and balancing the environment (through several parameters such as State of Charge (SOC), State of Health (SOH), State of Power, State of Safety, temperature, voltage, current and coolant flow). The BMS not only manages the functions of a battery actively but also estimates the status of the battery and predicts several crucial variables such as SOH, Remaining useful time (RUL), SOC, etc.

First-hand, we will be reviewing and discussing battery ageing mechanisms and data-driven technologies proposed behind the health estimation and lifetime predictions for Li-ion batteries (Li et al. 2019). Li describes the health management unit of a battery consisting of two major steps—the SOH estimation (estimating the extent of failure) which is then followed by the health prediction (predicting the battery performance variation and identifying the time of failure). The main degradation of a Li-ion can be categorized into 3 main modes—the loss of lithium inventory (LLI), the loss of active material (LAM) in the electrodes, and the increase of cell internal resistance. These 3 modes of battery degradation are based on 9 ageing mechanisms which are then influenced by 7 factors based on temperature variations, pressure variations, SOC/Voltage, current rates and time, etc.

SOH estimation: Some of the most faithful data-driven SOH estimation approaches are based on Differential analysis (DA) and Machine learning (ML). While DA involves feature identification from differentiated curves of mechanical, thermal, or electrical parameters during battery cycling, ML methods train a model in accordance with the extracted input features to describe the cell ageing behaviour to estimate the SOH.

Battery health prognostics: Health prediction includes Remaining Useful time (RUL) prediction and Capacity forecasting. RUL prediction is done by pre-defining a failure threshold signal and then the modelled degradation signal is sent to compare it with the threshold value. RUL is obtained by

$$\textbf{RUL} = \textbf{estimated life of the training units} - \textbf{current life position of the test units}$$

Similarly, capacity forecasting is done by predicting future changes in SOH as a function of battery usage history. Both the above-mentioned frameworks for battery health prognostics are based on two methods namely—Analytical models (requires the development of an ageing model) and ML methods (model-free and learn from the ageing data to predict battery health).

The accuracy of RUL prediction can further be improved by utilizing a big data analytics framework that combines any one of the EV battery modelling methods as discussed previously with the driving pattern analysis (Karmavijaya et al. 2019). After collecting internal resistance and voltage data, the driving data of each trip is extracted from EV cloud platform and transformed into a consistent vector for pattern recognition. For analysis, an unsupervised clustering approach Growing Hierarchical Self-Organizing Maps (GHSOM) is utilized to cluster driving patterns and thus the analysis of energy consumption of each driving pattern is done for automakers to regulate the power consumption estimations according to different driving patterns. Finally, on the basis of route plans uploaded by EV users and SOH estimates, the system calculates RUL and advises the users through the user interface.

Table 3 Data collection nodes in the proposed methodology (Rahimi-Eichi and Chow 2014)

Data nodes		
Weather	Temperature—to adjust battery parameters	Wind speed and direction—to estimate aerodynamic power consumption due to opposite direction flowing winds
Route and Terrain	Points of decision (turns) for the driver	Latitude and longitude for each point available in polyline data format
Battery Manufacturer	Resistances, Capacitances, currents, SOCs and several more parameters for battery modelling	
EV manufacturer	Velocity of EV, mass, slope, acceleration, friction, and several other parameters to calculate EV power consumption—for EV modelling	
Driver History	Speed and location of EV from the GPS history of driver	

To solve the driving range anxieties in users, (Rahimi-Eichi and Chow 2014) proposes a range estimation framework to pre-inform users the maximum possible driving range, collecting data from five collecting nodes described in Table 3.

Cell to cell variations are known to accelerate the decay rate of battery life and hence reducing the performance and safety of an EV battery. Based on previous research, (Lu et al. 2020) found the five indicators determining cell to cell variations—time variation, voltage variation, current variation, resistance variation, and SOC variation. The first step involves collecting voltage, time, current, and temperature from charging cloud data. The five indicators are then estimated using obtained original signals. This paper suggests a weighted scoring mechanism to score the cell-to-cell variations of the battery pack. While the scoring mechanism for all five indicators are established based on thresholds and weight factors, the thresholds are determined on real applications and the weights are determined on an analytics hierarchy process. Thresholds contribute to individual unweighted scores whereas weight factors contribute to weighted scores, and together they both contribute to the total final score. Finally, 3 battery packs are evaluated using the proposed method, and it is observed that the pack with the highest score has less ageing whereas the pack with the least score has deep ageing.

Infrastructural Aspect—Using trajectory data analysis, (Li et al. 2015) proposes an Optimal Charging Station Deployment (OCSD) framework which takes a historical EV taxi trajectory data, road map data, and existing charging station information as input, and performs Optimal Charging Station Placement (OCSP) and Optimal Charging Point Assignment (OCPA). Both the mentioned frameworks are designed to optimize the Charging Pile Network (CPN) of the city by setting up 'K' new charging stations in the city, with 'M' new charging points distributed over the set of K new charging stations. In simple terms, these two optimization components are designed to limit the average time taken by an e-taxi to travel to charging station, and the average waiting time for an available charging point in the city.

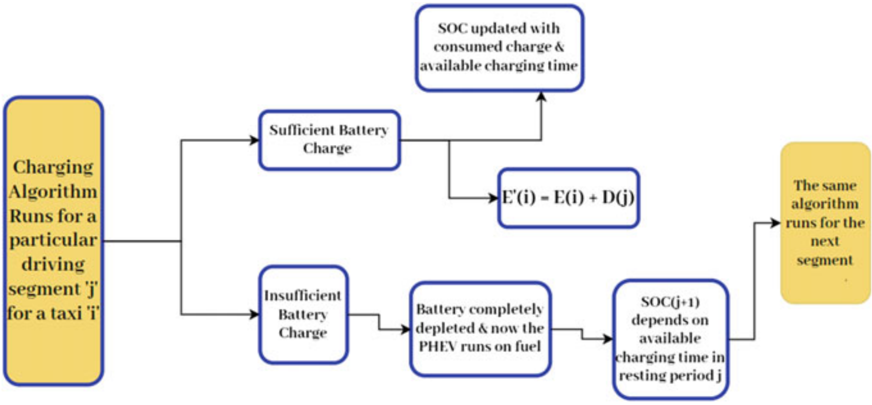


Fig. 10 Segment-wise charging algorithm proposed. Where, E(k) = electrified mileage of taxi 'k'. D(r) = total distance driven in segment 'r'. E'(k) = updated electrified mileage of taxi 'k'. SOC(r) = SOC at the beginning of segment 'r'

Trajectory data analysis is also utilized by (Cai et al. 2013) to characterize individual travel patterns by examining trajectory data of 10,375 taxis of Beijing by big data mining and further evaluates the impact of adopting plug-in hybrid EVs in the taxi fleet on GHG emissions, based on characterization of individual travel patterns. A Plug-in Hybrid EV charging algorithm for each segment is proposed in this study to analyse whether the total charge present in the battery is enough for the whole driving segment based on several inputs such as driving segments, resting periods, SOCs, and battery size. The charging algorithm is described in Fig. 10. Once all segments of a trip are analysed, the portion of the trip that can be electrified using a PHEV (with a given battery and charging conditions) is calculated. Next, several other factors are considered for simulations and modelling such as fuel cost saving and electrification rate, adoption, government subsidy, and GHG emissions to evaluate the extent to which each factor affects the PHEV adoption and lifecycle GHG emissions since lifecycle GHG emissions are dependent on the power grid, vehicle cycle, and the fuel cycle.

So far, it feels safe to assume that government subsidies could decrease lifecycle GHG emissions, however, that is not the case. With increased government subsidies, citizens are encouraged to adopt PHEVs and hence the electrification rate also increases. The emissions reduced by PHEVs are not enough to compensate or reduce the lifecycle GHG emissions. Hence, it can be concluded that lifecycle GHG emissions can only be reduced by switching to less carbon-intensive power grids (Cai et al. 2013).

An important step in aiming to popularize EVs is to analyse the current trends related to popularity. The most preferred charging times, the most preferred days, the most preferred locations, and how the charging behaviour is dependent on several other factors such as weather, holidays, etc. Wang et al. (2019) attempts to analyse the charging behaviour of users of a city utilizing the big data by making use of data mining technologies to explore the big data and further analyse the charging behaviour of users. Two different algorithms are used namely—the K-means algorithm and the Apriori algorithm. The first approach utilized K-means algorithm to classify EV users based on their charging habits to compare users with a huge difference in charging behaviours. In the second approach, the Apriori algorithm is used for association rules mining to explore charging behaviour of users according to weather conditions such as rain, snow, temperature, etc., and special days such as weekends, national holidays, public holidays. Analysis of EV charging also involves forecasting the charging demand for the future.

Furthermore, (Arias et al. 2016) proposes a forecasting model to predict EV charging demand based on historical real-word traffic distribution data and weather data. The technical architecture of the proposed model has four layers each for a specific target. The first two layers of data sources and storage extract data from local disks (data sources) on the computer and store them using the MATLAB store function. The stored data is provided in chunks to the third layer of data management which uses the MapReduce function of MATLAB for data handling. The fourth layer of data processing includes a cluster analysis as well as relational analysis for the identification of factors influencing traffic flow. Finally, a decision tree that classifies cars and buses was established using MATLAB. For charging demand forecasting, the user is expected to provide number of EVs, month and day as inputs to the model. The second step is to determine the weather and day type of the forecasting day from historical data sets. With these two forms of inputs, clusters and charging times (based on corresponding clusters) can be easily determined. A random sampling of data helps in determining the initial SOC. Finally, the EV charging demand can be calculated for the residential and commercial sectors.

Charging demand and Occupancy rates also play an important role in influencing the decision-making process of firms and agencies while determining potential locations of setting up new EV charging stations. Further examination of charging demand and occupancy rates give insights on electricity demand in particular regions. Lee et al. (2017) examines the charging demand in Jeju city utilizing open data sets of chargers across the city and open software components by performing a stream analysis to examine the number of occupied chargers over the city for as much as 500-h duration and increasing. Lee et al. (2017) also plots charging behaviour of the most frequent chargers and occupancy rates in different regions across the city.

3.6 Traffic Flow Predictions

Road transportation is the backbone of smart cities; it costs 1.35 million deaths (World Health Organisation 2020) and trillions of dollars to the global economy annually, damaging public health and the environment. The objective of traffic flow prediction is to provide traffic flow information that improves traffic efficiency, helps users to make better travel decisions, alleviate traffic congestion, reduce carbon emissions, etc. (Lv et al. 2014). Although traffic constitutes to be an integral part of a city, not much literature based on applications of big data-driven technologies in traffic exists. We begin our study by reviewing the sources of traffic data and then discussing the existing traffic flow prediction models.

Lv [31] lists the data sources related to traffic flow prediction as various sensor sources that include inductive loops, radars, cameras, mobile GPS (Geographical Positioning System), crowdsourcing, etc. Inductive loops are electromagnetic vehicle detection systems (Wikipedia 2021a), highly helpful in estimating the vehicle density on roads and the number of vehicles passing through a point location. The authors have thus proposed a deep-learning-based traffic flow prediction method, where a Stacked Auto-Encoder (SAE) model is used to learn generic traffic flow features, and it is trained in a layer-wise greedy fashion. Unlike the previously proposed ARIMA models and its variants that consider only the shallow structure of traffic data, the proposed model by (Lv et al. 2014) can successfully discover the latent traffic flow feature representation such as the non-linear spatial and temporal correlations.

Similar to traffic flow predictions, traffic congestion predictions also play an important role in providing better alternative routes to the drivers, thus reducing the time per trip and emissions as well as preventing future congestions. Two types of Long Short-Term Memory models (LSTMs) for congestion prediction based on data collected in a 5-min window are developed by (Majumdar et al. 2020). Model 1 is a univariate model based on the observed speed of the previous 5 min while Model 2 is a multivariate model based not only on previously observed speeds but also on traffic flow rates and vehicle headways. Since LSTM models are a subclass of Recurrent Neural Networks (RNNs) their training performances decrease with increasing data size and hence this study is primarily focussed on data analytics. In future, attempts can be made to utilize the big data generated using similar Machine Learning models for better congestion forecasting.

3.7 Smart Manufacturing

The manufacturing and production industry play a huge role in a company's economy, energy usage, and waste production. Almost all goods consume energy even before their work-life starts, as every manufacturing/production process needs energy. Similarly, the end of a product's economic life doesn't mean the end of its environmental impact. Hence, manufacturing methods need to be re-designed to use minimal energy and less environmental impacts. As laid down by (O'Donovan et al. 2015) smart manufacturing intends to convert data obtained over the product lifecycle into manufacturing intelligence so as to achieve positive impacts in all phases of manufacturing. Manufacturing frameworks are encountering an unstable growth of more than 1000 EB annually (Yin and Kaynak 2015) Hence, big data can prove to be an upgrade in the adequacy of smart manufacturing and there are several ways to go about it. Although not much literature exists to quantify the effects of utilizing manufacturing big data, we will be reviewing two pieces of literature that aim to utilize the manufacturing big data (1) to develop intelligent frameworks and streamline the manufacturing process (2) explore the relationship between big data and sustainable manufacturing.

Tao et al. (2018) proposes a very basic data-driven smart manufacturing framework where data from the manufacturing process is combined with data from orders and production plans. Intelligent algorithms and predictive models examine this data with the assistance of big data analytics so as to streamline the manufacturing process. The framework consists of four modules each for four such functions—input manufacturing data, data-driven analysis for information processing and generating recommendations, real-time monitoring to help encourage the efficient running of various types of equipments and problem processing by human operators or AI applications for a holistic approach.

Dubey et al. (2015) in their study aim to explore the relationship between big data and world-class sustainable manufacturing (WCSM). The methodology adopted consisted of identifying the constructs which impact upon sustainable manufacturing followed by principal component analysis (PCA) on the data collected. Constructs are formative and reflective and are studied by big data analytics perspective (measures, sampling design, data collection). Hence, the theoretical framework was generated, and it was tested using the gathered data.

3.8 Renewable Power Generation and Forecasting

An electric power generation system requires the adjustment of the generating and the loading components to be genuine and dependable and to operate safely at quality standards. Unfortunately, during the generation of power through RES, many issues are faced due to the fluctuations. Generation from RES results in uncertainty and variability to a high extent. The reason being changes in weather conditions due to which the intensity of ocean waves, winds, and sun rays might vary with time. Hence, it is important to develop frameworks supporting renewable power generation through a holistic approach. In this capacity, big data aids renewable energy power generation in mainly three capacities

- *Leveraging Renewable Energy (RE) Sources*: Newer power generating plants are now considering offshore renewables as the next potential source of clean energy after solar, and hence (Amaro and Pina 2017) specifically studied wave power as the target source to develop the largest maritime database through a multi-stage platform which is expected to integrate data of different velocity, variety, and volume under a trusted engine and finally produce a big data informative archive of value and veracity back to the participants and local communities.
- *Managing and analysing the performance of RE Systems*: Escobedo et al. (2017) designed a six-layer big data infrastructure for the management of PV Systems. The architecture of this management system is expected to perform general tasks of data ingestion, processing, storage, analysis, and monitoring of information in real time keeping in mind specific requirements of parallel processing, scalable

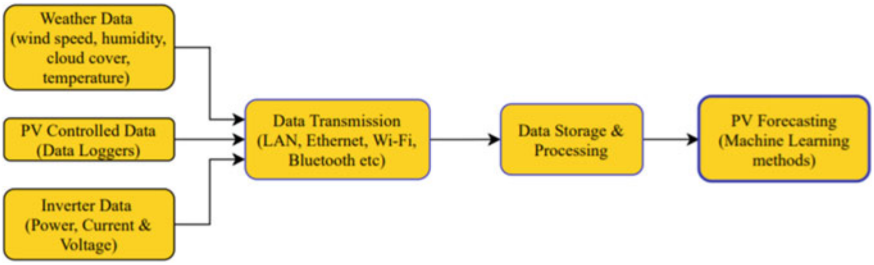


Fig. 11 Flowchart of PV forecasting (Preda et al. 2018)

infrastructure (horizontal, vertical, and fault-tolerant), and distributed storage. Additionally, the authors also perform a case study of power solar generation where a photovoltaic system is connected to a grid to estimate the working performances based on 22 essential factors (directly acquired by the collected data).

- *Forecasting the output renewable power*: Preda et al. (2018) proposed a methodology based on sensor applications where big data is captured from sensors and loggers, broken down and finally PV forecasting is done with the help of Support Vector Machine and Linear Regression algorithms (Fig. 11). By utilizing more parameters in the machine learning process, the root mean square error was observed to improve by considerable amounts.

4 Summary

Recently, a vast majority of the literature has contributed to research addressing challenges of sustainable development through big data-driven technologies and hence, the role of big data in several components of a smart city is pronounced, especially in the EV sector. One end that can be drawn from the examination of such big data aided methodologies is that embracing ICTs dependent on big data analytics and distributed computing storage facilities may support cooperation and correspondence between the various components of smart cities. With the growth and popularity of data-driven technologies, a significant amount of attention also goes to the advancements in the sector of data collection techniques, thus resulting in the growth of data collection devices such as AMIs, GIS, IEDs, sensors, actuators, etc. and finally all these combined result in the increased volume and variety of data. Integrating such huge amounts of heterogeneous data likewise assumes a significant job and a few arrangements have come up for the same, for example, virtual item and metadata. However, in any case, the source of data collected varies in each aspect. Similarly, big data tools (e.g. Hadoop, Cassandra, Hive) and techniques (Machine learning, Deep Learning, etc.) of big data have also evolved with time to address all kinds of data, matching data sets with the most suitable type of analysis models and visualization methods. Processing of data additionally varies in the few models delivered. While

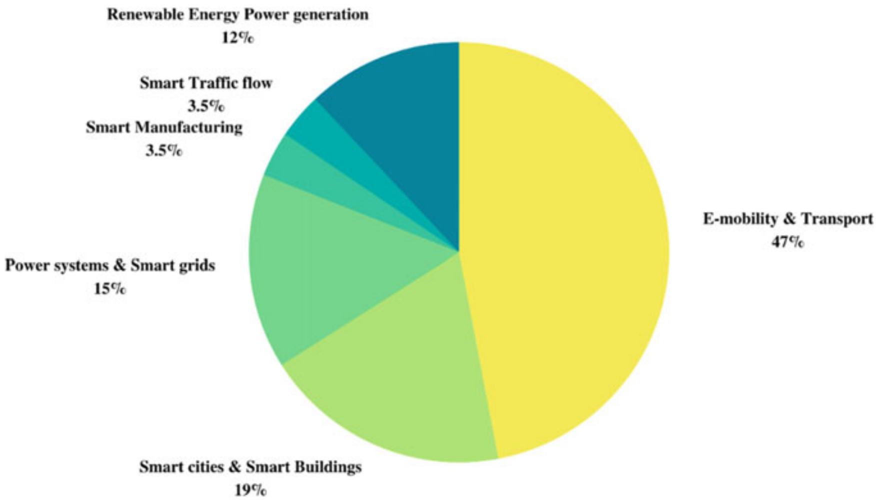


Fig. 12 Approximate quantitative analysis of the existing literature

all the above-mentioned frameworks use different data-driven technologies, they all aim to achieve the same goals in terms of sustainability with similar flow processes.

A common observation from the existing literature is that a significant amount of literature has been published for applications of big data in the transport sector, specifically towards E-mobility. Similarly, a considerable share of literature samples also focuses on the smart building and smart grid components of smart cities. A comparatively smaller amount of literature samples also discusses the forecasting and prediction of RES power generation as well as their monitoring and management systems. Smart manufacturing, traffic management, etc. have very limited research work at present (Fig. 12).

Concluding, Big data analytics sure has come up far in the digitalized era to ease sustainability-related challenges. However, big data techniques and methodologies need to be checked for suitability for the type of data, the application, and the desired knowledge to be obtained for their proper implementation. Some exploration holes despite everything exist in social and economic issues as well as in the proper implementation of technologies and there is an extension for research in those ways.

5 Limitations of Existing Research

The existing research surely does contribute a developing knowledge into existing theory, combines technologies and presents several new ideas in which the world can make use of the huge amounts of data generated to achieve goals of sustainability. However, some potential limitations shall be noted to overcome them in future research.

- The effect estimates in most of the models and frameworks presented above are just prospective observational studies and the results might differ in practical application, majorly due to inconsistency between the actual data and the procured data.
- IoV has a huge potential in creating more sustainable research opportunities due to the huge amounts of data generated. However, currently, very few IoV implementations are researched & several aspects such as safety, navigation, a social network on wheels, customer interests, etc. can be implemented. Safety issues also need to be addressed in network protocols (Xu et al. 2018).
- Few challenges associated with data-driven health estimation and prediction methods are
- ML methods are known to give less in-depth information regarding battery ageing & hence shall be combined with physical methods of battery decay such as DA to help in the identification of most sensitive indicators.
- Li-ion cells are also known to be environment sensitive hence compromising on the accuracy of models.
- Most of the existing research is done on the cellular level and the results might differ at a battery level.
- The sample size of the data collected needs to be increased and we need a more reliable method of analysing unstructured data (if collected by different methods) while enabling WCSM through big data techniques. Additionally, research needs to be done on data reduction techniques which would have helped to identify more enablers of WCSM (Dubey et al. 2015).
- Considering the evolution of sensors and loggers, a large number of PV systems exist without sensors. Hence, we are missing out on monitoring information while forecasting PV output. Even if the PV arrays are provided with data collection units, the accuracy of the collected data isn't guaranteed. For e.g. in Preda et al. (2018) the weather parameters are collected from weather stations for PV forecast which do not usually coincide with the location of PV array. Thus, the parameters are not accurate.
- While solar energy and wind energy are popular renewable sources of energy, ocean waves are still unpopular due to the several technological challenges associated with waves and hence extensive research is required in these fields. Some of the potential drawbacks of wave energy listed by Amaro and Pina (2017) are (1) Wave Energy Converters need to be designed to withstand extreme weather conditions (where the power may exceed over 100 times its mean value. (2) Energy storage means are required for the conversion of energy from slow random oscillations into useful motions to drive electric generators.
- Along with advantages in the application of big data, several issues need to be considered, such as the development of models that will give better accurate estimations for parameters in almost all climatic conditions and the development of reliable models for better prediction of weather data.
- Several confidential data sets and user data sets flow from one organization to the other in the process of procuring and analysing data. Several data protection steps need to be taken for the flow of such confidential data. In case of leaking of data,

it would be of crucial importance to identify from where the data got leaked & who was responsible for data leakage.

- The current literature doesn't discuss other monetary and social issues prompting atmosphere changes and contamination. Apart from smart cities and technical aspects of climate degradation, tourism is one such social issue. Tourism is heavily responsible for air, water, and land pollution due to stress on a particular land or water resource. Other factors may be due to more than usual traffic congestions and local shops becoming more expensive, stocks running out of shops and more generation of waste. These problems require predictive analysis and detailed forecasts utilizing big data techniques.
- Waste generation minimization and management are also important practices to achieve goals of sustainability. Big data analytics can not only help in the production of cleaner wastes (having less impact on the environment) but also help in the reduction of the amount of waste generated and manage them. Although no such study exists yet, there is scope for big data analytics and machine learning methods in the fields of waste segregation and management from homes to dumping grounds.
- Overall, most research articles aimed at achieving environmental sustainability through big data analytics indirectly aim to gain improvements in the business sectors as well. However, there is hardly any research addressing the social and economic aspects of sustainability directly. Although environmental sustainability sounds like a bigger challenge and rounds up the umbrella of sustainable development, the other two aspects can't be ignored since all three come hand in hand.

6 Scope of Future Work

Big data in itself is not a technology or a science field. Big data just enables us to get an actual idea of the real scenario taking place in the system to understand it better i.e. forming a basis for making technological and economic advancements. A better understanding of patterns, correlations, trends, etc. and deeper insights into the existing issues can then help in putting forward the technologies to overcome the issues. With the fast improvement of sensor technology, wireless transmission technology, network communication technology, cloud computing, distributed computing, smartphones (Zhou et al. 2016), etc. large amounts of data is generated every second. Previously, it was estimated that by 2020 1.7 MB data will be created every second for every person on earth by 2020 (Ahmad 2018). This data needs to be put to proper use for several issues of the modern world such as crisis response, disaster management and developing of smart management frameworks for transport systems, building systems, power systems, traffic systems, etc.

- Big data analytics aims to achieve the goals of clean power generation, efficient power transmission, dynamic power distribution, rational electricity consumption and innovative storage solutions by using one or many of the various data analysis techniques and algorithms, including optimization, forecasting, classification, and clustering, can be applied on the large amounts of smart grid big data (Sagiroglu et al. 2016)

- The efficient usage of building energy plays a vital role in the reduction of prices, minimizing harmful environmental effects, and brings ease for the users in terms of automation.
- The most predominant feature of this era is high-frequency generation and exchange of a large volume of data in real time that must be analysed to make proper decisions and to take appropriate actions as quickly as possible.
- Advances of big data in transport will help in large scale deployment of Battery Electric vehicles (BEVs) and Hybrid Electric Vehicles (HEVs) with de-carbonization as the main objective.
- Apart from the impact of GHG emissions on climate change, transport is also a significant source of noxious air pollutants. The data hence collected regarding these can help in improving the air quality for public health.
- Users are helped make better transport decisions by predicting traffic flow and monitoring traffic crowding by the huge data collected from various sensor sources, including inductive loops, radars, cameras, mobile Global Positioning, systems, crowdsourcing, social media (Lv et al. 2014).
- Heterogeneous network environment such as an industry or a service provider where delay in delivery of any action or goods will result in loss of the whole project it is necessary to have a system which can help to keep the record of item's location, current condition, and combining these details with inventory for faster delivery. Industrial IoT based on Big data analytics (IIoT) can be considered here (Lee et al. 2017). The same system can be used in a smart building where a user can wash clothes in a washing machine from anywhere just by connecting the washing machine to the internet. Of course, this technology is not just limited to washing machines but can help in all kinds of household chores. Thus, revolutionizing the concepts of household chores for several women across the world especially in developing Asian countries as well as for easing the tasks for students and bachelors across the world. However, the odds of achieving such a spread of machine intelligence in the near future is still debatable due to several risk factors and unreliability.
- As reviewed in several literature pieces, big data analytics plays an important role in the growth and development of an EV ecosystem for large scale deployment of EVs. Several technical data related to battery ageing and battery health play an important role in the development of efficient and cost-effective batteries, which solves major issues related to affordability, range anxieties, and charging time anxieties among users. Similarly, charging big data can be utilized for development in the market and infrastructural aspects of an EV ecosystem. Figure 13 represents a flow chart highlighting the potential usage of charging big data in the EV market and infrastructure.

- Big data can leverage the development of offshore renewables, particularly wave energy conversion systems. Potential locations for wave power development

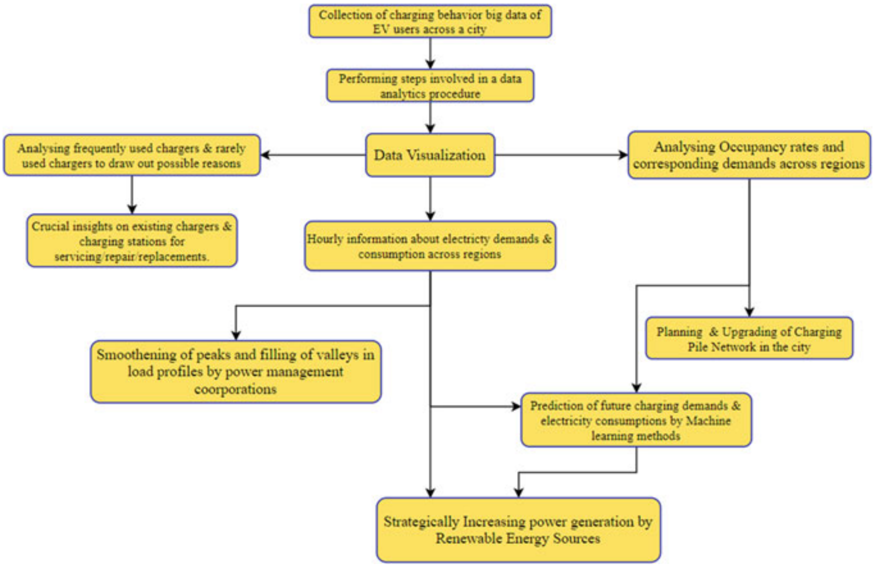


Fig. 13 Flowchart representing the usage of an EV user charging behaviour big data in future planning of several aspects of an EV ecosystem

in India are coasts of Maharashtra, Goa, Karnataka, and Kerala (Kanyakumari has the highest power owing to strong winds) (Wikipedia 2021b). On a whole, renewable energy generation systems are known to have several drawbacks limiting their popularity. Uncertainty of the source and less efficiency of the power generating system top the list. However, solar energy, wind energy, wave energy, and other forms of RES are infinitely available. Machine learning models based on big data proves to be the biggest achiever in this case for forecasting and prediction.

- Government records and surveys are an incredible source of basic information about each and every citizen and hence, the government agencies can utilize the citizen data for the identification of the needy to provide them with the necessary assistance. From the business point of view, it is highly expected that an adjustment in the viewpoint at business issues through big data analytics will provide a wide scope of modifications for a wide range of associations and supply chains, in turn helping their business grow. Business owners can leverage the big data generated to understand their customers better and improving business operations answering both environmental and economic concerns. Since organizations need to hire a special team of data engineers, scientists, and analysts to assess the real case scenario of a business, the introduction of big data analytics also generates new income opportunities. Hence, through the business sector, big data can be expected to have a hand in dissolving issues related to the economic sustainability of a country on a larger scale.

7 Case Study

7.1 Test System

From the EV charging station placement point of view, it is crucial to monitor and hence realize the power supply demands pressure in a region and the subsequent heavy loading on distribution transformer due to both—residential charging stations and commercial charging stations. However, in this study, we focus mainly on the analysis of charging demand and occupancy of commercial charging stations because residential chargers are very unlikely to happen for any densely populated Indian city in near future. The proposed strategy not only helps in relieving the burden on distribution transformer but also in decreasing the cost of EV charging by making it more economical.

The purpose of a case study is to provide new information, trends, and results to EV charging station management firms by the analysis and examination of traffic and charging big data sets (Fig. 3). Thus, we perform a case study by assuming charging demands in the city of Mumbai, India based on route traffic congestion analysis with reference to the studies performed in Li et al. (2015) and Wang et al. (2019). The methodology used in this case study is described in Fig. 14.

The development of an EV ecosystem and deployment of EVs are both cause and effect of each other. Due to limited on road EVs in Mumbai and due to less popularity among Indian users as well as charging station management firms there exists a lack of sensors and other data collection devices at EV charging stations.

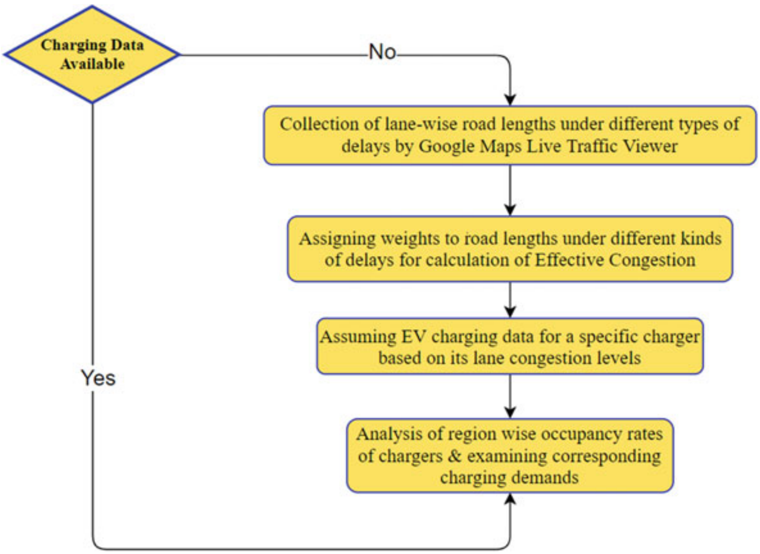


Fig. 14 Route Congestion analysis methodology utilized in the case study

Hence, the historical time stamp data of chargers is unavailable. With the same background reasoning, we attempted to collect the remaining data related to the EV infrastructure for Mumbai.

- Mumbai suburban and Mumbai city combined have around only 27 eV charging stations till date.
- Most of the charging stations are equipped with one or two chargers at max while data regarding number of chargers is not available for the rest.
- Types of chargers available are mainly AC slow chargers and DC fast chargers.

7.2 Route Selection

Most of the EV charging stations can be categorized under two major expressways in Mumbai—Eastern Expressway and the Western Expressway, hence Set of n routes $R = \{R_1, R_2\}$ where $n = 2$. Part 1 of Table 4 gives additional first-hand information about features of both routes considered in our study. Out of 27 charging stations, 3 of them either do not come under any route or are a part of Navi-Mumbai and hence, are excluded from our analysis. Figure 15 shows the Map of Mumbai with locations of charging stations situated along its major express highways.

7.3 Selection of Time Frames

- Charger occupancy observational time frames are divided mainly into rush hours and off-peak travel hours. The biggest reason behind these rush hours is that almost all organizations have similar work timings. Similarly, off-peak travel hours are named so as they have comparatively lesser traffic congestion due to no reasonable cause.
- Traffic congestion data unlike charger status data changes at every minute. However, changes at extremely small intervals are negligible. Thus, traffic congestion data from Google live traffic viewer is collected for the previously stated 7-time frames of a day with shorter one-hour intervals.

Part 2 of Table 4 represents traffic congestion and charger occupancy observational selected time frames for analysis.

7.4 Route Congestion Analysis

Congestion road length data for a week's time, i.e. 3 August 2020 (Monday)–9 August 2020 (Sunday) was collected from Google Maps Live traffic viewer. The

Table 4 Represents important set of information considered in the performed case study

Part 1—Information of Selected Routes		
Features	Eastern express highway (R_1)	Western express highway (R_2)
Number of charging stations	8 + 2 (Shared with R_2)	14 + 2 (Shared with R_1)
Route length under observation	17.4 kms	30.5 kms
Start—Destination	Parel to Vikhroli	Borivali to Bandra + Worli Sea link
Part of Highway	National Highway—3	National Highway—8
Part 2—Time Frames		
Charger occupancy observational time frames		Traffic congestion observational time frames
Rush Hours	Off-Peak Travel Hours	
7–9 AM	1 AM – 4 PM	7–8 AM
		8–9 AM
		1–2 PM
5–7 PM	8–9PM	3–4 PM
		5–6 PM
		6–7 PM
		8–9 PM
Part 3—Route Lengths highlighted under different kinds of delay in Google maps		
Colour of Highlighted part	Congestion caused delay	
Green	No delay	
Orange	Slight delay (S)	
Red	Medium delay (M)	
Maroon	Heavy delay (H)	

road lengths are highlighted under different kinds of delay on the Google Maps live Traffic viewer as represented in Part 3 of Table 4.

- *Lane Considerations*—The total distance highlighted under each colour is measured and numerically added for each time frame. Lane-wise congestion analysis plays an important role in determining charging behaviour as an EV user is likely to in a particular direction via lane 1 is much more likely to search and spot for an EV charging station lying adjacent to Lane 1 than the ones lying adjacent to Lane 2A route generally comprises of two lanes owing to vehicles flowing in the opposite directions.
- *Assigning weights*—Proper weights are assigned to each road length depending on the type of delay it has. Assuming the speed of the vehicle to be ‘x’ for heavy delay in one of the many road types, the vehicle speeds for medium and slight

Fig. 15 Charging stations represented in Blue lie under Route 1 whereas stations in Red lie under Route 2. Orange highlighted stations lie equidistant between both routes whereas the charging stations in black are situated in Navi Mumbai or do not lie in the area surrounding any route

delays can be assumed to be '2x' and '3x', respectively for the same road type. Through lane-wise congestions of both routes we derive the graph in Fig. 16 which represents the effective congestion of both routes after assimilating all kinds of delays by assignment of weights (Figs. 15 and 16).

Therefore, effective congestion in metres $= \frac{S}{3} + \frac{M}{2} + \frac{H}{1}$

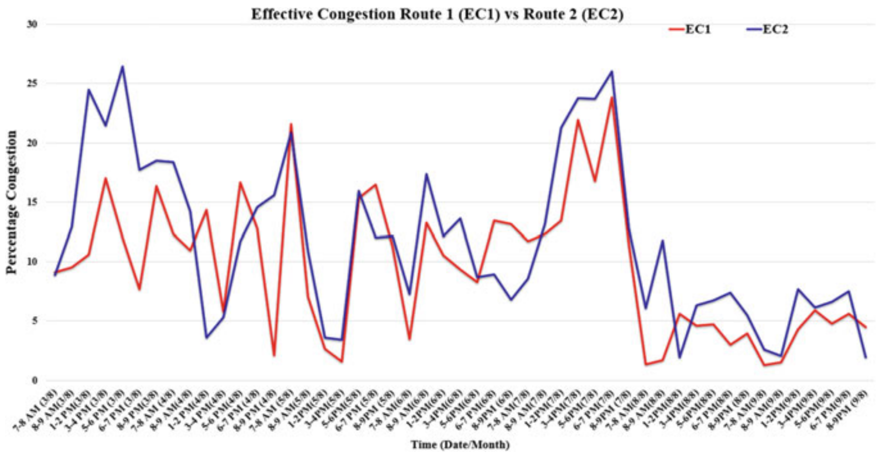


Fig. 16 Comparison of route congestions of route 1 and route 2

7.5 Charging Data Assumption

Occupancy of chargers across the city at a particular time frame was assumed on the basis of several constructive reasons adding up, lane congestion levels being the strongest of them all since the number of vehicles on road is directly proportional to the charging demand especially in case of EVs majorly due to the need of frequent charging requirements. EVs require frequent charging due to many reasons

- EVs need to be charged before the battery reaches below 20% to avoid cycles of deep discharge which adds up to the limited storage capacities of batteries.
- Charging less than its capacity—To avoid damage to the battery due to generated heat. Further, research suggests that charging an EV battery to its full capacity can decrease the range of the EV (InsideEVs n.d.).
- Lesser driving range—As compared to an ICE car, most EV batteries can store charge capable of running much lesser than ICE cars due to the battery capacities.

Also, different EV users can be expected to charge their EVs differently, some might prefer charging for shorter intervals regularly while others might charge for full capacity at once less frequently.

7.6 Route -Wise Charging Data Analysis

With the previously explained basis of charging data assumption, charging data is assumed. Further, plots are generated so as to identify the peaks, valleys, and points of comparison (Fig. 17).

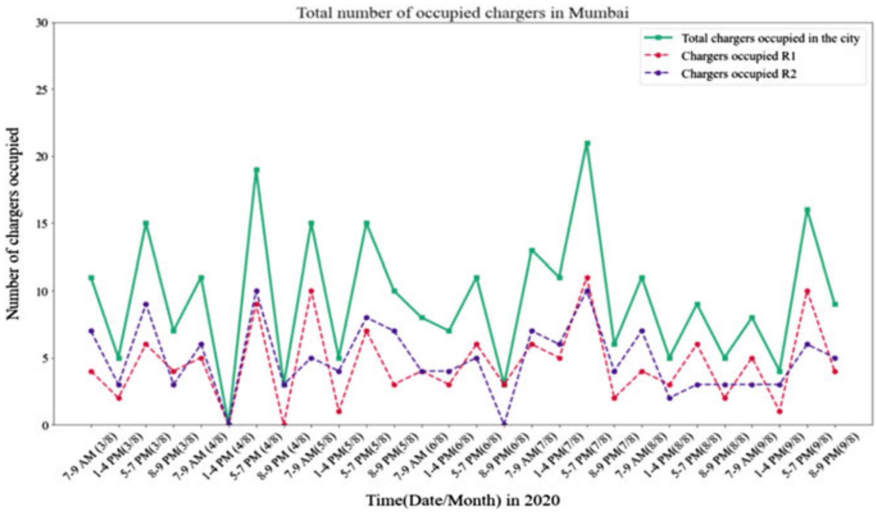


Fig. 17 Represents the number of chargers occupied on a whole for both routes and total chargers occupied in the city

7.7 Region Wise Charging Data Analysis

Mumbai mainly comprises of two districts of Maharashtra—Mumbai Suburban and Mumbai City. Mumbai Suburban is further divided into three sub districts/regions—Andheri, Borivali, and Kurla as in Fig. 19a. Thus, we have four major regions of importance—Borivali, Andheri, Kurla, and City. Table 5 describes the spread of EV charging stations and chargers across the city.

‘Charging Demand’ and ‘Occupancy Rates’ are two terms which are often used interchangeably. However, there is a slight difference in between both the terms which also accounts for the logical relation between them. The relation is crucial from business point of view as it helps firms with deeper insights while aiming to optimize a charging pile network in a city. The relation can be explained by Fig. 18.

Bar graphs are plotted with the assumed charging data to examine occupancy rates in each of the four districts/regions of Mumbai. Figure 19b is a bar graph representation of percentage occupancies at each time frame for each region. However, it doesn’t deliver a clear picture of overall occupancies of regions in order to make out which region has the most occupancy or charging demand. Figure 19c represents effective region wise occupancy of a week for all regions.

Table 5 Region wise allocation of charging stations and chargers

	Borivali	Andheri	Kurla	City
Charging stations	3	9	5	7
Chargers	4	13	10	9

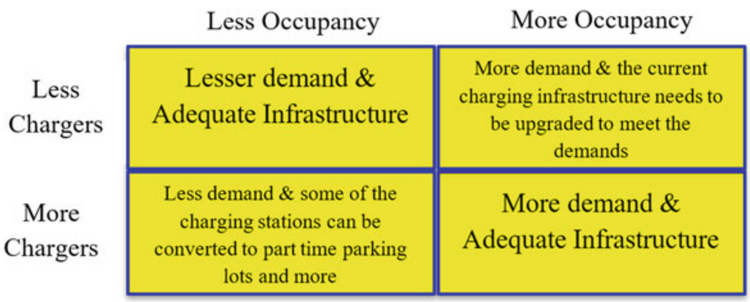


Fig. 18 Relationship between demand and occupancy

We intend to draw out conclusions utilizing the above proposed relations from the generated graphs. It would be safe to assume that Borivali region draws out most of the EV users, while it still has very few chargers. Hence, it is safe to conclude that the charging infrastructure in Borivali needs to be upgraded to meet the future charging demands. Similarly, Andheri represents a situation close to Case 2. Even after having the maximum number of chargers in Andheri, occupancy rates remain the lowest. And hence, it is important for firms to analyse the least occupied charging stations to convert them into part time parking lots or for other purposes.

7.8 Predication of Congestion Data—Without Pandemic Conditions

Real time traffic congestion levels for Mumbai is collected from the live traffic viewer website of *TomTom* (www.tomtom.com n.d.) for the same time frame during the year 2019 to assess the conditional changes without the effect of pandemic. Further, this traffic data is not comparable to the traffic congestion levels previously calculated for Pandemic conditions and is made comparable by converting time-based congestion into distance-based congestion.

7.9 Route-Wise Congestion Analysis for Without Pandemic Conditions

Similar graphs as in the previous case of Pandemic conditions are plotted for normalized conditions of the city, to compare the distance-based congestions of pandemic and without pandemic conditions and thus make predictions about the charging demand. Figure 19 is an overall representation of congestion data comparison of 2019 versus 2020 which is approximated to suit our congestion observation time frames. The peaks and valleys of congestion line plots are incident at the same time

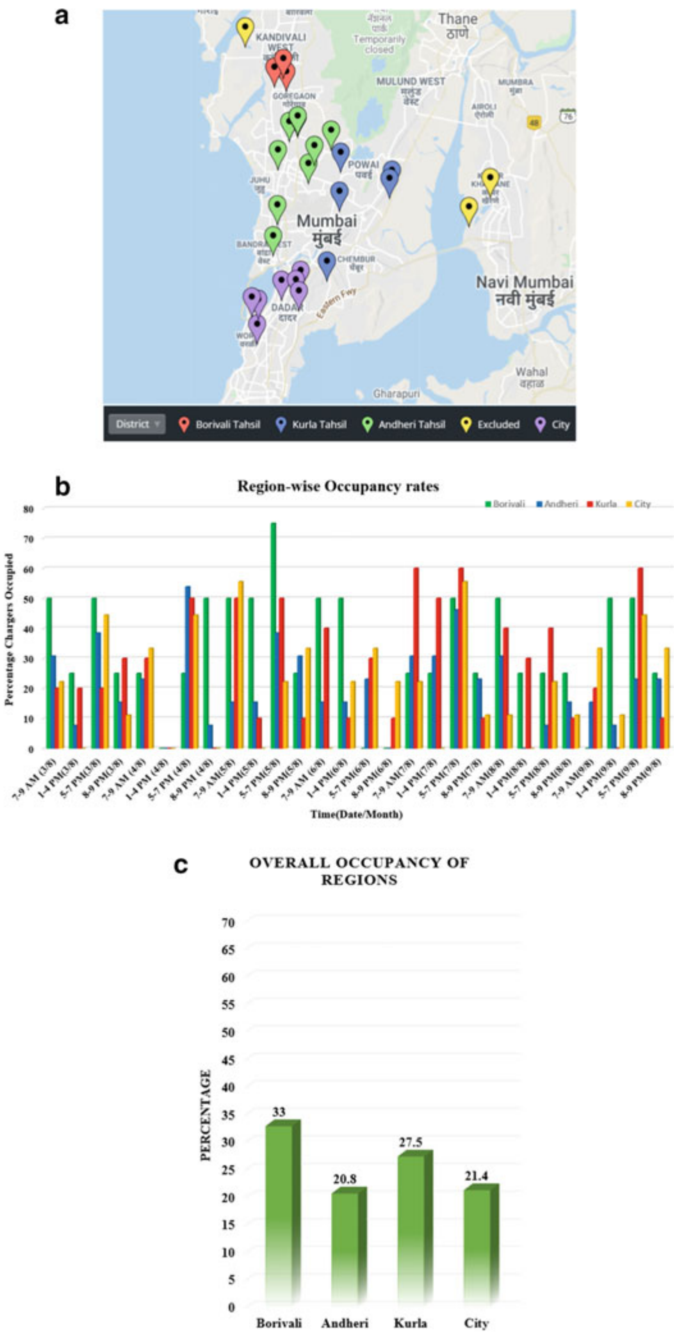


Fig. 19 **a** District-wise division of charging stations in Mumbai. **b** Region wise Occupancy rates of chargers in Mumbai. **c** Overall percentage occupancy of regions over a week

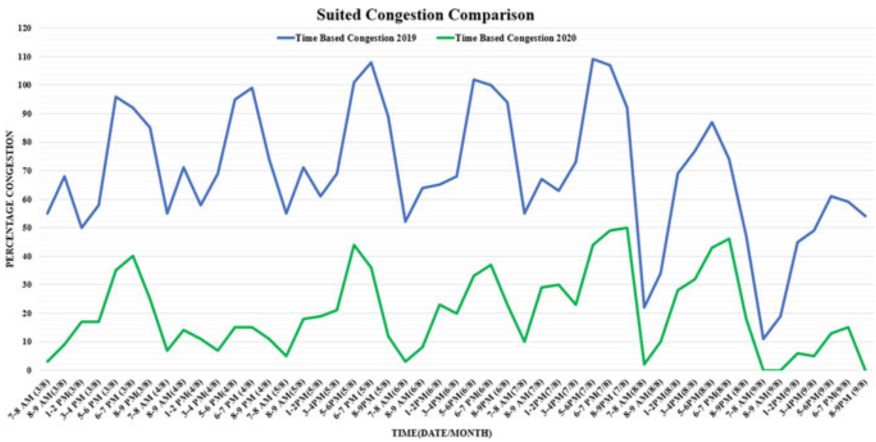


Fig. 20 Overall congestion comparison for the city of Mumbai by the obtained congestion data from TomTom N.V

frames in Fig. 20, thus, suggesting to us the travel patterns of users (rush hours and off-peak hours) remain the same, but differ in magnitude for both cases.

7.10 Charging Data Assumption—Without Pandemic Conditions

With the predicted congestion rates of 2020 without pandemic conditions, charging behaviour was assumed and plotted for both lanes of both routes as well as for all four regions. Further, Fig. 21a, b compare the occupancy rates of the without pandemic conditions with the pandemic conditions.

Figure 21b suggests to us similar charging occupancy ranks of regions in ‘without pandemic’ conditions as in the ‘pandemic conditions’ since the factors determining charging behaviour in a particular region remain the same, and only the amount of EV users seeking charging increases. However, in ‘without pandemic’ conditions Kurla region seems to have highest occupancy rates followed by Borivali. Hence, it is safe to conclude that the region of Kurla might come under Case 2, i.e. the current infrastructure is suitable for demands. Also, a considerable increase in the number of occupied chargers in the city resulted in slight region wise occupancies, which suggests to us the denounced effects of overall city occupancy rate in region wise occupancies. Further, more visual graphs need to be examined in a similar way and cases need to be identified for different regions in a city for the successful deployment of Charging Pile Network in the city.

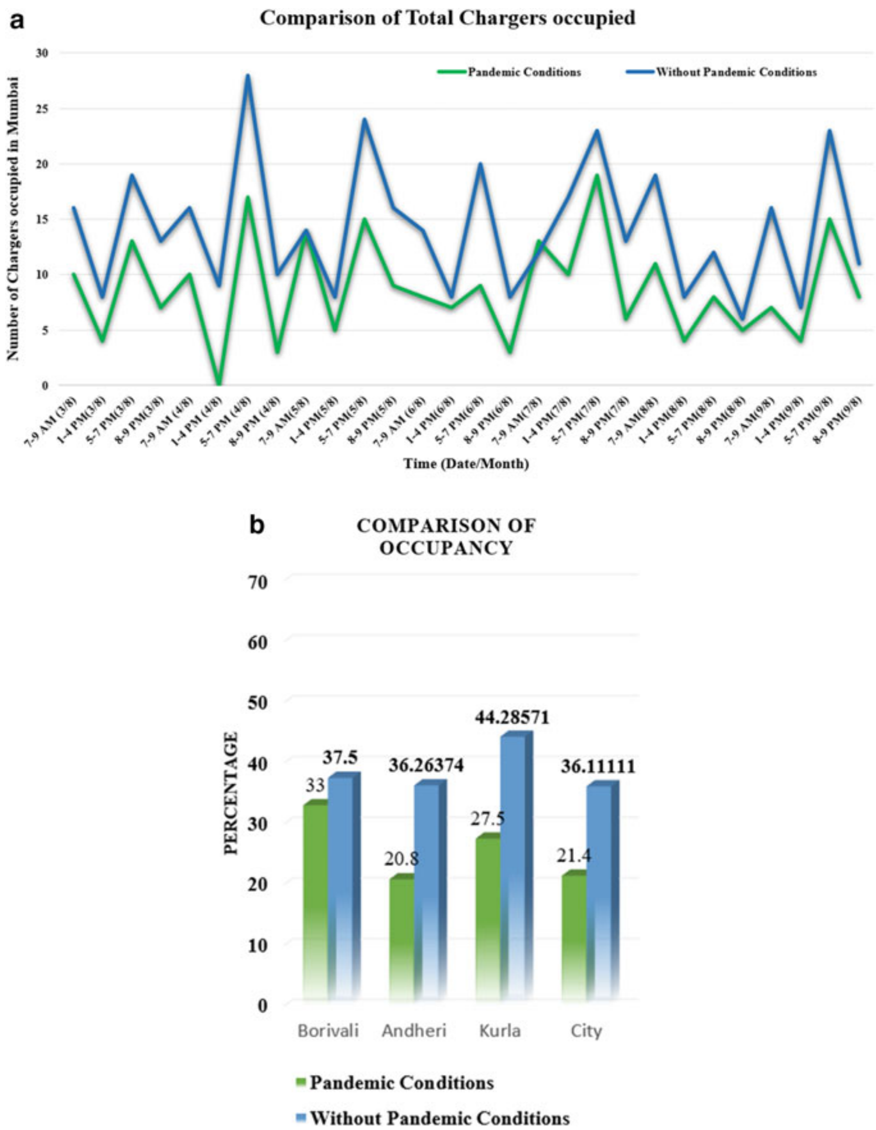


Fig. 21 **a** Comparison of total number of chargers occupied during pandemic conditions and without pandemic conditions. **b** Comparison of Region wise overall percentage occupancies in Pandemic conditions versus Without Pandemic Conditions

8 Discussion and Conclusion

Big data is produced at a fast rate every second through a huge number of data generating and recording sources. With the increasing population, one can only imagine the amount of data generated in the coming years. This chapter aims at providing a state-of the-art literature review of the recent research advances in achieving goals of sustainable development through big data-driven technologies majorly through components of a smart city. Although the goal is to cover all three aspects of sustainable development, this chapter largely focuses on achieving goals of environmental sustainability mainly because of two reasons—(1) Most existing research articles focus on environmental sustainability-driven business and economic viewpoints (maximizing profit through the utilization of renewable sources and implementation of environmentally sustainable practices (2) We believe environmental sustainability is a prerequisite; opening doors for citizens, organizations, and governments to achieve socio-economic aspects of sustainable development. Having said that, the need for research in economic and business viewpoints need to be examined in depth for a holistic picture of the data being generated.

The case study performed revolves around the 'most-raw utilization of data'—analytics. Examining and exploring the data to influence the decision-making tasks of organizations is one of the most beneficial and wise uses of data sets generated. With the same mindset, we attempt to analyse the current Charging Pile Network in the city of Mumbai. The E-Mobility sector is expected to grow exponentially in India once the EV technologies address the cost and time anxieties for better user preferences. Thus, it is important to understand the current charging infrastructure, study the electricity demands, and monitor the EV sales in order to prepare the city for mass adoption of EVs.

With the necessary background, readers are encouraged to research focussing on overcoming one of the many limitations of the existing literature articles thus help in increasing the domain of reach of data-driven sustainable development. We also hope our research opens new research directions for researchers to explore the usage and potential of data-driven technologies for a wider range of tasks including analytics, optimization, forecasting, management, risk analysis, location of faults, and many more in the future.

References

Ahmad I (2018) How much data is generated every minute? [Infographic]. Social Media Today. https://www.socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/

Amaro N, Pina JM (2017) Big data in power systems leveraging grid optimization and wave energy integration. In: 2017 international conference on engineering, technology and innovation (ICE/ITMC). Funchal, pp 1046–1054. https://doi.org/10.1109/ICE.2017.8279997

Ancillotti E, Bruno R, Conti M (2013) The role of communication systems in smart grids: architectures, technical solutions and research challenges. Comput Commun 36:1665–1697. https://doi.org/10.1016/j.comcom.2013.09.004
Apache.org (2019) Apache Hadoop. http://hadoop.apache.org/
Arias M, Bae S (2016) Electric vehicle charging demand forecasting model based on big data technologies. Appl Energy 183.https://doi.org/10.1016/j.apenergy.2016.08.080
BloombergQuint (n.d.) Electric vehicle sales in India up 20% in 2019–20, industry body says. https://www.bloombergquint.com/business/electric-vehicle-sales-in-india-up-20-in-2019-20-industry-body-says. Accessed 31 May 2021
Cai H, Xu M (2013) Greenhouse gas implications of fleet electrification based on big data-informed individual travel patterns. Environ Sci Technol 47.https://doi.org/10.1021/es401008f
Chesterton A (2018) How many cars are there in the world? CarsGuide. https://www.carsguide.com.au/car-advice/how-many-cars-are-there-in-the-world-70629
Chou J-S, Ngo N-T, Chong WK, Gibson GE (2016) Big data analytics and cloud computing for sustainable building energy efficiency. In: Start-up creation, pp 397–412
Daissaoui A, Boulmakoul A, Karim L, Lbath A (2020) IoT and big data analytics for smart buildings: a survey. Procedia Comput Sci 170:161–168
Datadog (2016) Hadoop architectural overview. https://www.datadoghq.com/blog/hadoop-architecture-overview/. Accessed 31 May 2021
De la Rue du Can S, Letschert V, Mcneil M, Zhou N, Sathaye J, Orlando E, Berkeley L (2009) Residential and transport energy use in India: past trend and future outlook. https://doi.org/10.2172/951788
Dittrich J, Quiané-Ruiz J-A (2012) Efficient big data processing in Hadoop MapReduce. In: Proceedings of the VLDB endowment, vol 5, no 12, pp 2014–2015
DoSomething.org (2021) 11 Facts about global poverty. https://www.dosomething.org/us/facts/11-facts-about-global-poverty#fnref7. Accessed 31 May 2021
Dubey R, Gunasekaran A, Childe SJ, Wamba SF, Papadopoulos T (2015) The impact of big data on world-class sustainable manufacturing. The Int J Adv Manuf Technol. 84(1–4):631–645. https://link.springer.com/article/https://doi.org/10.1007/s00170-015-7674-1
Dutton WH, Blumler JG, Kraemer KL (1987) Wired cities: shaping future communication. Macmillan, New York
EDF Energy (2015) EV: Batteries. EDF Energy. https://www.edfenergy.com/electric-cars/batteries
Escobedo G, Jacome N, Arroyo-Figueroa G (2017) Big data & analytics to support the renewable energy integration of smart grids—case study: power solar generation. In: Proceedings of the 2nd international conference on internet of things, big data and security
Gennaro M, Paffumi E, Martini G (2018) Big data for supporting low-carbon road transport policies in Europe: applications, challenges and opportunities", case studies on transport policy 6(4)
Goos G, Hartmanis J, Van Leeuwen J, Ishida T, Isbister K (2000) Digital cities. Heidelberg Springer, Berlin Heidelberg, Berlin
Green Car Reports (n.d.) What Is EVSE and why does your electric car charger need it? https://www.greencarreports.com/news/1050948_what-is-evse-and-why-does-your-electric-car-charger-need-it. Accessed 31 May 2021
Growth of electricity sector in India from 1947–2020 government of India ministry of power central electricity authority New Delhi (2020). https://cea.nic.in/wp-content/uploads/pdm/2020/12/growth_2020.pdf
Guardknox (n.d.) The EV charging ecosystem. https://www.guardknox.com/electric-vehicle-charging-ecosystem/. Accessed 31 May 2021
Indian consumers: second place. (n.d.) https://www.nationalgeographic.com/greendex/assets/dl/pdf/Indians.pdf. Accessed 31 May 2021
InsideEVs (n.d.) You should set EV Charge limit to just 60 percent: here's why. https://insideevs.com/news/368097/video-60-percent-ev-charge-limit-benefits/. Accessed 31 May 2021

Jakubiec JA, Reinhart CF (2012) A method for predicting city-wide electric production from photovoltaic panels based on LiDAR and GIS data combined with hourly DAYSIM simulations. Solar Energy 93(C):127–43

Johnson J (2021) Global digital population 2019|Statistic. Statista. https://www.statista.com/statistics/617136/digital-population-worldwide/

Junaidi N, Shaaban M (2018) Big data applications in electric energy systems. In: 2018 international conference on computational approach in smart systems design and applications (ICASSDA)

Karmawijaya MI, Nashirul Haq I, Leksono E, Widyotriatmo A (2019) Development of big data analytics platform for electric vehicle battery management system. In: 2019 6th international conference on electric vehicular technology (ICEVT)

Khalid MR, Alam MS, Sarwar A, Jamil Asghar MS, A comprehensive review on electric vehicles charging infrastructures and their impacts on power-quality of the utility grid. eTransportation. https://doi.org/10.1016/j.etran.2019.100006

Khan M, Babar M, Ahmed SH, Shah SC, Han K (2017) Smart city designing and planning based on big data analytics. Sustain Cities Soc 35:271–279

Kitchin R (2013) The real-time city? Big Data and Smart Urbanism. GeoJournal 79(1):1–14

Komninos N (2002) Intelligent cities: innovation. Routledge, Knowledge systems and digital spaces

Kuhlman T, Farrington J (2010) What is sustainability? Sustainability 2(11):3436–3448

Lee J, Park G-L, Han Y, Yoo S (2017) Big data analysis for an electric vehicle charging infrastructure using open data and software. In: Proceedings of the eighth international conference on future energy systems

Li Y, Luo J, Chow C-Y, Chan K-L, Ding Y, Zhang F (2015) Growing the charging station network for electric vehicles with trajectory data analytics. In: 2015 IEEE 31st international conference on data engineering

Li Y, Liu K, Foley AM, Zülke A, Berecibar M, Nanini-Maury E, Van Mierlo J, Hoster HE (2019) "Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. Renew Sustain Energy Rev, Elsevier, 113(C):1–1

Liang F, Yu W, An D, Yang Q, Fu X, Zhao W (2018) A survey on big data market: pricing, trading and protection. IEEE Access 6:15132–15154

Linder L, Vionnet D, Bacher J-P, Hennebert J (2017) Big building data—a big data platform for smart buildings. Energy Procedia. 122:589–594. https://doi.org/10.1016/j.egypro.2017.07.354

López-Ibarra JA, Goitia-Zabaleta N, Herrera VI, Gazta ñaga H, Camblong H (2020) Battery aging conscious intelligent energy management strategy and sensitivity analysis of the critical factors for plug-in hybrid electric buses. eTransportation 5:100061

Lu Y, Li K, Han X, Feng X, Chu Z, Lu L, Huang P, Zhang Z, Zhang Y, Yin F, Wang X, Dai F, Ouyang M, Zheng Y (2020) A method of cell-to-cell variation evaluation for battery packs in electric vehicles with charging cloud data. eTransportation 6:100077

Lv Y, Duan Y, Kang W, Li Z, Wang F-Y (2014) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 1–9

Majumdar S, Subhani MM, Roullier B, Anjum A, Zhu R (2020) Congestion prediction for smart sustainable cities using IoT and machine learning approaches. Sustain Cities Soc 102500

Miller H (2011) Collaborative mobility: using geographic information science to cultivate cooperative transportation systems. Procedia—Soc Behavioral Sci 21:24–28. https://doi.org/10.1016/j.sbspro.2011.07.005

Months TG 9, Weeks 4 (2020) Terabytes of data were buried in the Arctic. We ask the person in charge, why? Mashable SEA. https://sea.mashable.com/tech/11810/terabytes-of-data-were-buried-in-the-artic-we-ask-the-person-in. Accessed 31 May 2021

Moradi J, Shahinzadeh H, Nafisi H, Marzband M, Gharehpetian GB (2019) Attributes of big data analytics for data-driven decision making in cyber-physical power systems. In: 2020 14th international conference on protection and automation of power systems (IPAPS)

Ndukwu M, Onwude D, Bennamoun L, Abam F, Simo-Tagne M, Horsfall I, Briggs T (2020) Nigeria's energy deficit: the challenges and eco- friendly approach in reducing the energy gap. Int J Sustain Eng 1–18

Nitesh Jain (2015) Big data and hadoop essentials. Udemy. https://www.udemy.com/course/big-data-and-hadoop-essentials-free-tutorial/. Accessed 6 Nov 2019
O'Donovan P, Leahy K, Bruton K, O'Sullivan DTJ (2015) An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. J Big Data 2(25):1–26
Onya (2021) Greendex ranks 18 countries based on sustainability. https://www.onyalife.com/greendex-ranks-18-countries-based-on-sustainability/. Accessed 31 May 2021
Preda S, Oprea S-V, Bâra A, Belciu (Velicanu) A (2018) PV forecasting using support vector machine learning in a big data analytics context. Symmetry 10(12):748
Rahimi-Eichi H, Chow M-Y (2014) Big-data framework for electric vehicle range estimation. In: Presented at the 40th annual conference of the IEEE industrial electronics society (iecon2014). IEEE, Dallas, TX
Rathore MM, Paul A, Hong W-H, Seo H, Awan I, Saeed S (2018) Exploiting IoT and big data analytics: defining smart digital city using real-time urban data. Sustain Cities Soc 40:600–610. https://www.sciencedirect.com/science/article/abs/pii/S2210670717309782
Sagiroglu S, Terzi, R, Canbay Y, Colak I (2016) Big data issues in smart grid systems. In: 2016 IEEE international conference on renewable energy research and applications (ICRERA)
Sanila S, Subramanian D, Sathyalakshmi S (2017) Real- time mining techniques: a big data perspective for a smart future. Indian J Sci Technol 10(42):1–7
Science (2018) Cities emit 60% more carbon than thought in embodied consumption, C40 report says. https://www.nationalgeographic.com/science/article/city-consumption-greenhouse-gases-carbon-c40-spd. Accessed 31 May 2021
Selecthub.com (2021) Big data versus business intelligence versus data mining|the differences. https://www.selecthub.com/business-intelligence/bi-vs-big-data-vs-data-mining/. Accessed 31 May 2021
Sharma G, Sood VK, Alam MS, Shariff SM (2020) Comparison of common DC and AC bus architectures for EV fast charging stations and impact on power quality. eTransportation 5:100066
Shepard M (2011) Sentient city: ubiquitous computing, architecture, and the future of urban space. MIT Press, Cambridge, Mass
Snowden S (2021) Melting of greenland's ice sheets has passed the point of no return, scientists say. Forbes. https://www.forbes.com/sites/scottsnowden/2020/08/14/greenlands-melting-ice-sheets/?sh=74244bba2449. Accessed 31 May 2021
Softwaretestinghelp.com (2019) Top 15 big data tools (big data analytics tools) in 2019. https://www.softwaretestinghelp.com/big-data-tools/
Solability (2021) The global sustainable competitiveness index. https://solability.com/the-global-sustainable-competitiveness-index/the-index. Accessed 31 May 2021
Tao F, Qi Q, Liu A, Kusiak A (2018) Data-driven smart manufacturing. J Manuf Syst 48:157–169
Team B (2021) First economic contraction in 4 decades: India's GDP shrinks 23.9% in Q1, FY21. Business-standard.com. https://www.business-standard.com/article/economy-policy/first-economic-contraction-in-4-decades-india-s-gdp-shrinks-x-in-q1-fy21-120083101022_1.html. Accessed 31 May 2021
The Conversation (2021) Siberia heatwave: why the Arctic is warming so much faster than the rest of the world. https://theconversation.com/siberia-heatwave-why-the-arctic-is-warming-so-much-faster-than-the-rest-of-the-world-141455. Accessed 31 May 2021
The Guardian (2021) Earth has lost 28 trillion tonnes of ice in less than 30 years. https://www.theguardian.com/environment/2020/aug/23/earth-lost-28-trillion-tonnes-ice-30-years-global-warming. Accessed 31 May 2021
The Hindu (2021) Data|An estimated 12.2 crore Indians lost their jobs during the coronavirus lockdown in April: CMIE. https://www.thehindu.com/data/data-over-12-crore-indians-lost-their-jobs-during-the-coronavirus-lockdown-in-april/article31520715.ece. Accessed 31 May 2021
The Indian Express (2020) Explained: How does the Rewa Solar Power Plant match up to similar plants in India and abroad? https://indianexpress.com/article/explained/rewa-solar-power-plant-madhya-pradesh. Accessed 31 May 2021

Tien J (2013) Big data: unleashing information. J Syst Sci Syst Eng 22(2):127–151

Udemy (n.d.) Free Hadoop tutorial—Hadoop starter kit. https://www.udemy.com/course/hadoopstarterkit/. Accessed 31 May 2021

Union of Concerned Scientists (2014) Car emissions and global warming. Union of concerned scientists. https://www.ucsusa.org/resources/car-emissions-global-warming

Vanani IR, Majidian S (2019) Literature review on big data analytics methods. Social media and machine learning. https://www.intechopen.com/books/social-media-and-machine-learning/literature-review-on-big-data-analytics-methods

Vikaspedia.in (2020) vikaspedia domains. https://vikaspedia.in/energy/energy-efficiency/electric-vehicles-in-india

Voivontas D, Assimacopoulos D, Mourelatos A, Corominas J (1998) Evaluation of renewable energy potential using a GIS decision support system. Renew Energy 13:333–344. https://doi.org/10.1016/S0960-1481(98)00006-8

Wang S, Moriarty P (2018) The potential for big data for urban sustainability. https://doi.org/10.1007/978-3-319-73610-5_3

Wang Y, Cai S, Shen Y, Shao D, Gong X, Zhou C, Chu D (2019) The analysis of electrical vehicles charging behavior based on charging big data 63–67. https://doi.org/10.1109/ICCCBDA.2019.8725776

web.archive.org (2018) Physical progress (achievements)|ministry of new and renewable energy|government of India. https://web.archive.org/web/20180503151432, http://mnre.gov.in/physical-progress-achievements. Accessed 31 May 2021

Wikipedia Contributors (2019a) Electric car use by country. Wikipedia https://en.wikipedia.org/wiki/Electric_car_use_by_country

Wikipedia Contributors (2019b) Motor vehicle. Wikipedia. https://en.wikipedia.org/wiki/Motor_vehicle

Wikipedia (2021a) Induction loop. https://en.wikipedia.org/wiki/Induction_loop

Wikipedia (2021b) Wave power in India. https://en.wikipedia.org/wiki/Wave_power_in_India. Accessed 31 May 2021

World Economic Forum (2021) The plastic pandemic is only getting worse during COVID-19. https://www.weforum.org/agenda/2020/07/plastic-waste-management-covid19-ppe/. Accessed 31 May 2021

World Health Organization (2020) Road traffic injuries. who.int. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Worldometers.info (2021) COVID live update: 171,029,617 cases and 3,556,677 deaths from the coronaviru—worldometer. https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?. Accessed 31 May 2021

Wu P-J, Chen Y-C (2017) Big data analytics for transport systems to achieve environmental sustainability. In: 2017 international conference on applied system innovation (ICASI)

www.power-technology.com (n.d.) The 100% club: the cities going all-out on renewables. https://www.power-technology.com/features/100-club-cities-going-renewables/. Accessed 31 May 2021

www.tdworld.com. (n.d.) StackPath. https://www.tdworld.com/electrification/article/20972652/transportation-electrification-how-to-do-it-right. Accessed 31 May 2021

www.tomtom.com (n.d.) Mumbai traffic report|TomTom Traffic Index. https://www.tomtom.com/en_gb/traffic-index/mumbai-traffic/

Xu W, Zhou H, Cheng N, Lyu F, Shi W, Chen J, Shen X (2018) Internet of vehicles in big data era. IEEE/CAA J Automatica Sinica 5(1):19–35

Yevgen B, Jinrong Q (2013) Battery power management for portable devices. Artech House, Cop, Boston

Yin S, Kaynak O (2015) Big data for modern industry: challenges and trends [Point of View]. In: Proceedings of the IEEE, vol 103, no 2, pp 143–146. https://ieeexplore.ieee.org/abstract/document/7067026/

Zhou K, Fu C, Yang S (2016) Big data driven smart energy management: From big data to big insights. Renew Sustain Energy Rev 56(January):215–225