doi:10.20944/preprints202204.0044.v1

Smart City Data Science: Towards Data-Driven Smart Cities with Open Research Issues

Iqbal H. Sarker[a,1,*]

[a] *Department of Computer Science and Engineering, Chittagong University of Engineering & Technology, Chittagong-4349, Bangladesh.*
[b] *Swinburne University of Technology, Melbourne, VIC-3122, Australia.*

Abstract

Cities are undergoing huge shifts in technology and operations in recent days, and 'data science' is driving the change in the current age of the Fourth Industrial Revolution (Industry 4.0 or 4IR). Extracting *insights or actionable knowledge* from city data and building a corresponding *data-driven model* is the key to making a city system automated and intelligent. Data science is typically the study and analysis of actual happenings with historical data using a variety of scientific methodology, machine learning techniques, processes, and systems. In this paper, we concentrate on and explore "Smart City Data Science", where city data collected from various sources like sensors and Internet-connected devices, is being mined for insights and hidden correlations to enhance *decision-making* processes and deliver better and more intelligent services to citizens. To achieve this goal, various *machine learning* analytical modeling can be employed to provide deeper knowledge about city data, which makes the computing process more actionable and intelligent in various real-world services of today's cities. Finally, we identify and highlight ten *open research issues* for future development and research in the context of data-driven smart cities. Overall, we aim to provide an insight into smart city data science conceptualization on a broad scale, which can be used as a reference guide for the researchers, professionals,

*Corresponding Author
Email address: `msarker@swin.edu.au` (Iqbal H. Sarker)
URL: `https://orcid.org/0000-0003-1740-5517` (Iqbal H. Sarker)

as well as policy-makers of a country, particularly, from the technological point of view.

Keywords: `Smart cities, data science, machine learning, Internet of Things, data-driven decision making, intelligent services, cybersecurity.`

1. Introduction

Nowadays, the world is experiencing an evolution of "Smart Cities" according to the needs of the people in the urban areas, which is typically considered as an application of the Internet of Things (IoT) [1]. IoT is defined as a connected network of heterogeneous components that are sensing, collecting, transmitting, and analyzing data for intelligent systems and services. With the advent of smart devices and their recent developments, the notion of linking everyday objects across existing networks has become highly favorable and is growing day by day. For example, in 2022, there will be about 20.4 billion linked stuff worldwide, while in 2020 it was around 8.4 billion linked stuff [2]. IoT has a huge impact on our lives in various dimensions such as social, commercial as well as economic. The IoT industry is projected to rise in revenue from 892 billion in 2018 to 4 trillion by 2025 in terms of expanding the digital economy [2]. Connected smart devices in the IoT network, can share and access authorized information to make intelligent decision-making. Thus, IoT provides essential building components for smart cities, i.e., data acquisition, data analytics, and intelligent decision-making as well as application handling for smart city services. This is shown graphically in Figure 1 highlighting two major components such as data-to-insights, and insights-to-applications towards "Data-driven smart cities", in which we are interested this paper.

Smart cities have risen to prominence in recent decades as a result of rapid urbanization around the world, notably in industrialized countries such as the United States, Australia, Canada, Europe, Singapore, Korea, Japan, and others. A smart city is typically an urban environment or an advanced modern city that

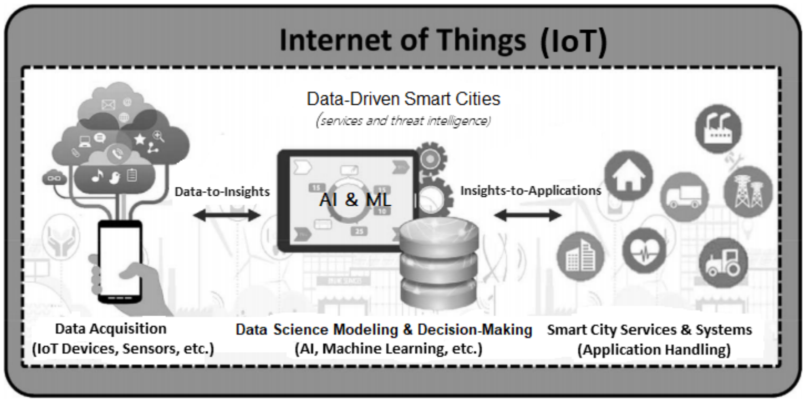

Figure 1: An understanding of data-to-insights and insights-to-applications towards 'Data-driven smart cities' within the environment of the Internet of Things (IoT).

ensures the availability of the city resources in terms of social, economical, and environmental aspects, to improve the quality of life of the citizens. According to the reports in [3] [4], the global urban population is expected to reach around 70% by 2050. The ongoing demand for city services due to such a huge amount of people and urbanization will have drastic impacts on cities' environment, proper management, and services, as well as cybersecurity. For instance, several security issues include illegal access to information, anomalies or attacks such as malware, botnet, denial-of-service (DoS), distributed denial-of-service (DDoS), social engineering, zero-day attacks, etc. [5] are considered as serious concerns about risks in the growth of smart environments and may cause disruptions in such smart services. Thus, to realize the ideal concept of a smart city, the cyber-threats and corresponding security policies are also needed to take into account for building smart cities.

A smart city modeling typically uses information and communication technologies (ICT) technologies to collect and share useful information, increase the efficiency of city operations, and improve the quality of services and citizens' life, which eventually leads to smart outcomes. City-data collected from diverse sources such as sensors, Internet-connected devices, or other external sources, is being mined for discovering useful insights and hidden correlations to provide

 doi:10.20944/preprints202204.0044.v1

better services to citizens and improve decision-making processes. Thus, the smart city services are supported by the technologies of the Fourth Industrial Revolution (4IR) [6] [7] like the Internet of Things (IoT), data analytics, artificial intelligence (AI), and machine learning that form the base of the data-driven modeling and can play a vital role to provide smart services as well as cybersecurity intelligence in smart cities utilizing the collected city-data. Therefore, the concept of "Data-Driven Smart Cities" could be an effective solution to manage and optimize the city resources and to provide smart services, such as smart people and governance, smart environment, smart transportation, smart health, smart grid, smart cybersecurity management and so on discussed briefly in Section 4. These smart services are highly promising to improve the quality of life of the people in cities in terms of automation, efficiency, and comfort.

In this paper, we explore "smart city data science", to establish data-driven smart cities that can address the difficulties of ongoing urbanization and increased population density in today's cities. As stated briefly in Section 3, analytical approaches such as machine learning and deep learning modeling are used to extract actionable insights or deeper knowledge about city data, making the computing process more actionable and intelligent. Overall, we aim to provide an insight into smart city data science conceptualization, thinking, modeling, and processing, where the applicability of machine learning approaches to data-driven intelligent decision making in smart city services has also been explored.

The contribution of this paper is summarized as follows:

- This paper mainly concentrates on and explores "smart city data science", where city data collecting from diverse sources such as sensors, Internet-connected devices, or other external sources, is being mined for insights and hidden correlations to enhance decision-making processes and deliver better and more intelligent services to citizens.
- We also explore machine learning-based analytical modeling that can be used to provide deeper knowledge about city data for data-driven smart

cities, which makes the computing process more actionable and intelligent.

- We then emphasize real-world smart city services on a broad scale and categorize them into ten potential work domains including smart people and governance, environments, transportation, energy management, public safety, cybersecurity management, etc. and explore how data-driven decision-making can play a key role in the context of smart cities.
- Finally, we identify and highlight ten open research issues within the scope of our study for future development and research in the context of data-driven smart cities.

The rest of the paper is organized as follows: Section 2 presents the background and reviews the related work within the area of smart cities. In Section 3, we briefly explore data science modeling including machine learning analytical model building. In Section 4 we have discussed broadly the smart city services with data-driven decision-making. In Section 5, we highlight the open research issues and future directions and finally, Section 6 concludes the paper.

2. Background and Related Work

In this section, we summarize and discuss the fundamental concepts of data-driven smart cities, background, and related work to highlight the scope of this study.

2.1. Defining Data-Driven Smart City

A "Smart City" typically uses ICT (information and communication technology) to boost economic growth, improve quality of life, and strengthen government systems. For example, a local authority might connect its transportation and energy grid systems, construct sensor-equipped energy-efficient buildings, and develop communications to improve healthcare, emergency, and other public service monitoring and access.

 doi:10.20944/preprints202204.0044.v1

In generic terms, a smart city is defined as "an urban environment that mainly enhances the performance and efficiency of the regular city operations and the quality of services provided to the citizens using ICT and other related technologies, as well as to achieve the economic growth". Formally, many experts have defined the term smart city considering various aspects and perspectives. For instance, a popular definition states that "a smart city connects physical, social, business and ICT infrastructure to uplift the intelligence of the city" [8]. In another comprehensive definition, "smart city is defined as an advanced modern city that utilizes ICT and other technologies to improve the quality of life, competitiveness, operational efficacy of urban services while ensuring the resource availability for present and future generations in terms of social, economic, and environmental aspects" [9].

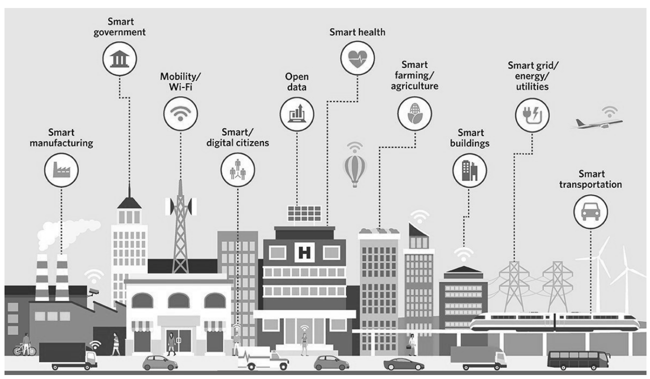

Figure 2: An overview of smart city components.

The primary objective of the original smart cities was to boost urban citizens' quality of life by reducing the contradiction between demand and supply in different functionalities [1]. However, modern smart cities especially focus on sustainable and efficient solutions in various sectors such as smart people and governance, energy management, transportation, smart building, health care, public safety, and many more, to meet the extreme necessities of urbanization. Figure 2 shows an overview of smart city components highlighting these application areas, discussed briefly with intelligent decision-making in a later section as well. Furthermore, smart cities or surroundings are increasingly vulnerable to various cyber-attacks or threats, like malware, denial-of-services, social engineering, zero-day attacks, and so on, as a result of technological advancement [5]. Thus, the concept of "Data-Driven Smart Cities" could be an effective solution to manage and optimize the city resources and provide smart services and security. The reason is that city data collected from various sources like sensors and Internet-connected devices, is being mined for insights and hidden correlations, to deliver intelligent and secured services to citizens and enhance decision-making processes. For effective data science modeling, various machine learning techniques can be utilized to achieve the goal, providing a deeper knowledge about data or actionable insights, making the computing process automated, smart, and intelligent. In the following, we highlight several key terminologies relevant to data-driven smart cities.

- *Data Science*: Within the scope of our study, we can define "data science" simply as the study of city data, where a data product will be a data-driven service. Thus, in the context of smart cities, it extracts useful knowledge or insights from city data for smart decision-making in various real-world applications domains [10].

- *Internet of Things (IoT)*: City data is collected and analyzed using IoT devices such as Internet-connected devices, sensors, lights, meters, and so on. The data is then used by the cities to improve infrastructure, public utilities, services, and other aspects of their operations [5].

- *Machine Learning*: Machine learning automates analytical model building through learning from the collected city data. Various machine learning techniques [11] can be used while building data-science models for smart cities, depending on the nature of the data.

- *Deep Learning*: Deep learning also automates analytical model building through learning from the collected city data. However, it uses multiple layers to progressively extract higher-level features from the raw city data.

Various deep learning techniques [12] can be used while building data-science models for smart cities, depending on the nature of the data.

- *Artificial Intelligence (AI)*: AI typically deals with giving machines the ability to seem like they have human intelligence [13]. In the context of smart cities, the above-mentioned machine and deep learning are considered AI techniques, which automate analytical modeling building using city data.
- *Context-Aware Computing*: This enables citizens' immediate needs to be anticipated and enriched, situation-aware, and usable content, services, and experiences to be offered proactively [14]. Thus the knowledge of data science can be used to extract insights from such contextual data. For instance, in our earlier paper, Sarker et al. [15] [16], we have discussed how different types of contextual information, such as temporal, spatial or locational, environmental, social, etc. impact differently in the relevant applications or services for the users.
- *Cybersecurity*: Due to technological advancement, smart cities are increasingly facing various cyber-attacks or threats. Thus the term cybersecurity is a concern, which is the practice of protecting critical systems and sensitive information from digital attacks, where data science modeling could be an effective solution [5].

The technological terminologies with their computing capability mentioned above could play a key role in various circumstances of the smart cities, depending on the problem nature and solution perspective.

2.2. Related work and the Scope of this Study

In recent years, the world is experiencing an evolution with the concept of "Smart Cities" according to the needs of the people in the urban areas. Several studies have been done on smart cities and related technologies. For instance, authors in [18] define the concept of the smart city. They also analyze the factors determining the performance of smart cities by performing statistical and graphical analyses. In [19], the authors present a structure that can be used to explore how local governments are envisioning smart city projects to grasp the idea of smart cities. The authors explain the definition of the word 'smart' in [20] in the sense of cities through an in-depth literature analysis of related studies as well as official documents from international institutions. The authors of [21] concentrate on how ICT is combined with existing urban infrastructures, coordinated, and integrated for the future of smart cities using emerging digital technologies. The authors concentrated on evaluating the discrepancies between sustainable and smart cities based on the number of indicators available in various frameworks in [22].

In addition to the conceptual discussion, the authors in [23] present the Internet of Things (IoT) for smart cities. They will explore a general reference system for urban IoT architecture, including basic features of urban IoT, as well as a web-based approach to IoT service design and related protocols and technologies. In [1], the authors discuss the features and characteristics, generic architecture, composition, and real-world implementations of smart cities. In [24], the authors analyze and discuss the key findings of the issues related to smart cities considering several aspects, such as smart mobility, living, government, etc. In [25], the authors focus on how city data is related to big data concerning its size, and thus discussed big data, smart cities, and city planning. The applications of big data to support smart cities have been presented in [26].

In [27] the authors addressed the role of advanced sensing in smart cities. The authors discuss smart cities as ecosystems of open and user-driven innovation for experimenting and validating potential Internet-enabled services in [28]. Different IoT-based machine learning mechanisms have been presented in [29]. IoT modules that are crucially needed and have changed the lives of the masses have also been emphasized. The authors explore different applications of artificial intelligence and machine learning in smart cities in [4] as well. The authors analyze different machine learning methods in [30], which address the challenges posed by IoT data by considering smart cities as the key use case. In

 doi:10.20944/preprints202204.0044.v1

[31], a systematic analysis has been presented on info mining and machine learning approaches for smart cities. An AI-based city analysis has been presented in [32], where the goal is to better understand the relationship between core AI technologies and their key application areas in urban planning and development, with an emphasis on Australian states and territories. In another study, [33] they explore whether artificially intelligent cities can protect humanity from natural disasters, pandemics, and other crises.

Unlike the previous studies, this article focuses on "Data-Driven Smart Cities" which can help cities deal with the challenges of ongoing urbanization and growing population density through data-driven intelligent decision-making and services in the context of smart cities. We aim to explore it using data science knowledge through machine learning-based analytical modeling, which is well known as a core technology in today's Fourth Industrial Revolution (Industry 4.0).

3. Smart City Data Science Modeling

In this section, we explore data science modeling for smart cities and its related components within the scope of our study. Thus this section includes three major parts; (i) understanding city data, (ii) data science process, and (iii) machine and deep learning-based analytical modeling. In the following, we briefly discuss these key parts according to the focus of this paper.

3.1. Understanding City Data

Data science is largely driven by the availability of data [10]. Thus, datasets, which are collections of data records with many properties or features and associated facts, are the foundation of smart city data science. Depending on the various potential reasons for city service growth, different cities focus on different domains. For example, 41% of the data sets from New York City [34] are related to city governance and education, with a focus on economic and educational development. Thus it's essential to understand the nature of smart city data that including the key elements for data-driven modeling.

 doi:10.20944/preprints202204.0044.v1

City-data is collected in a variety of sectors, including transportation, emergency services, public safety, public health, social media, the environment, city planning, etc. Transportation data includes things like geographic information, traffic mobility history, public transportation performance, and traffic anomalies. For instance, car parks, cycling lanes, petrol stations, bus stops, and electric vehicle charging stations are some examples of Barcelona transportation data [35]. Furthermore, several cities, including Beijing [36], Chicago [37], London [38], regularly publish data on various modes of transportation, including taxis, buses, urban trains, and bicycles. The level of safety provided to city residents, the level of preparedness of the city's emergency services, and the recurrence of emergencies, criminal activity, and natural disasters are all included in emergency and public safety statistics. For example, timestamps, location, incident type, and the text of the people's emergency alert are all included in the emergency notification data sets from New York City [34]. Similarly, energy data, of New York [34] focuses on energy usage and production in the city, including electricity, water, gas, oil, solar energy, and other public and private energy resources, that can be utilized to save wasted energy, optimize city energy distribution, and provide customized energy services through assessing consumers' energy usage habits. The majority of city environment data is concerned with the environment or surroundings in which people live or work. Sensor data collection techniques and services have been integrated into social networks, and social data is connected with human behaviors and social interactions. For example, Kosala and Adi [39] gather real-time road traffic information from Twitter timelines for real-time mapping in Jakarta, Indonesia.

These data are of great value in getting a greater understanding of city mobility, producing new services, and enhancing city performance. In recent years, governments, city departments, institutes, researchers, and people have gathered and shared city data across a wide range of sectors. For example, the US government's open data portals now have over 183,500 data sets of American cities, with an average of 2,791 new data sets added per month [40]. According to the data characteristics and target city services, data-driven smart cities are

 doi:10.20944/preprints202204.0044.v1

built on extracted insight from such types of relevant datasets using data science modeling, discussed in the following section.

3.2. *Data Science Modeling*

In this section, we briefly explore how data science can play a crucial role in data-driven smart cities. Figure 3 depicts a data science modeling process that begins with real-world city-data collecting and ends with data-driven services and automation. Each module of the data science process is briefly discussed in the context of smart cities in the subsequent subsections.

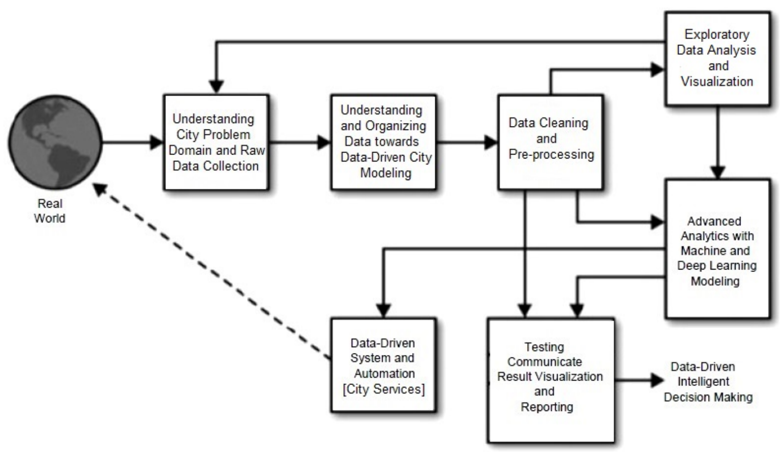

Figure 3: An example of data science modeling from real-world city data collection to intelligent data-driven services and decision-making in the context of smart cities.

- *Understanding City Problem Domain and Raw Data Collection:* There are several applications in various smart city domains, including environment, transportation, health, energy, public safety, and so on, discussed briefly in Section 4. In a particular domain, analyzing and providing appropriate answers to questions such as how much or many, which category or group, is the behavior or activities unrealistic or abnormal, which option should be chosen to take, what action should be taken, and so on may assist in gaining a better understanding of the requirements for a particular city issue and what we should extract from data. Raw city data from various data sources mentioned earlier can be collected once the problem has been identified. For example, sensed IoT data, the social media network, electronic medical records, daily crime logs, etc. are some examples of collected data in the domain of smart cities. The amount of data generated by society, city infrastructure, and digital technology is incredible today and will continue to increase dramatically in the future. The proliferation of IoT technologies, sensors, and cloud platforms has accelerated the generation process.

- *Understanding and Organizing Data towards Data-Driven City Modeling:* Since data science is mostly driven by the availability of data, a thorough understanding of the data is essential before constructing a data-driven model or system [10]. Data analysis employs analytical and logical reasoning to extract information from data. The main purpose of data analysis is to find meaning in data so that one can make better decisions with the information gathered. Thus, in the field of data science, interpreting the raw data collected for a certain problem domain, as indicated above, is important. Several factors need to be considered to clearly understand the data for a specific city problem, such as data type or format, data quantity, whether the data is sufficient or not, authorized access, feature importance, metrics to report the data, and so on. Overall, this module includes specifying which data is best suited for a data-driven city model and how to organize it.

- *Data Cleaning and Pre-processing:* The raw city dataset is transformed into a comprehensible format during data pre-processing [6]. Any analytic algorithm's results are directly influenced by data pre-processing methods. Data cleaning is the process of removing or fixing incorrect, corrupted, wrongly formatted, duplicate, or incomplete data from a dataset. When combining multiple data sources, there are numerous opportunities for data duplication or mislabeling. The data preparation technique, which often includes cleaning and transforming raw data before processing and analysis, is crucial for assuring data quality. The reason for this is that real-world data sets are typically noisy, have missing values, inconsistencies, or other data issues that are needed to be resolved [41]. The correct data or data quality must be sourced and cleansed to gain actionable insights, which is a requirement for any data science application.

- *Exploratory Data Analysis and Visualization:* Exploratory data analysis is the process of describing data using statistical and visualization tools to highlight relevant features of the data for further investigation [42]. This method analyses a large data set in an unstructured manner to uncover initial trends, attributes, points of interest, and so on to generate relevant data summaries. Thus data exploration is commonly used to determine the essence of data and to generate a preliminary assessment of its quality, quantity, and characteristics. A statistical model can be employed or not, but it mostly provides tools for developing hypotheses by visualizing and interpreting data using graphical representations such as charts, plots, and histograms [43] [44]. Before the acquired city data can be used for modeling, it's required to employ data summary and visualization to audit the quality of the data and offer the information needed to process it.

- *Advanced Analytics with Machine and Deep learning Modeling:* Data scientists create a model or a group of models to handle the specific city problem after the data is processed for modeling. Model creation is reliant on the sort of analytics required to address the problem, such as anticipating traffic jams under various traffic conditions, disease diagnosis, or predicting various types of security threats, such as malware attacks. Different forms of the machine and deep learning models can be developed to achieve the goal of best fitting the data according to the type of analytics, discussed briefly in Section 3.3. Various model validation and assessment metrics, such as accuracy, error rate, false positive, false negative, precision, recall, f-score, ROC, applicability analysis, and so on, are used to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0044.v1

measure model performance [10] [45]. Furthermore, sophisticated analytics such as feature engineering, algorithm tuning, ensemble methods, or building new algorithms can be used by machine learning professionals or data scientists to improve the ultimate data-driven model for tackling a specific city problem through smart decision-making.

- *Intelligent Decision-Making, Automation and City Services:* This is the final stage of the data science process, which could be intelligent decision-making, or services in the context of smart cities, also known as data products. A city data product is a system or tool that uses city data to assist individuals or businesses in making intelligent decisions or services. Non-data scientists can utilize data science to provide predictive analytics, descriptive data modeling, data mining, machine learning, risk management, and a variety of analysis methodologies through data products with a friendly user interface. The best data products can assist businesses and organizations in extracting information from their data to make predictions, reduce costs, and generate more revenue. For instance, corporations can use the outcomes of data science modeling to gain helpful information such as time series prediction and customer segmentation and then use that information to make better decisions.

Overall, we can conclude that data science modeling can be employed to promote improvements in smart city services. Data scientists analyze and manage data to find answers to big questions that help city authorities make better decisions and resolve problems. In summary, a 'smart city data scientist' explores city data to acquire a better knowledge of how a specific business or system runs, and then designs machine learning analytical models or tools centered on data-driven advanced analytics to aid decision-making. In other words, a 'smart city data scientist' is in responsible of processing, modeling, analyzing, and drawing conclusions from city data in order to improve the intelligence and smartness of city services.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0044.v1

3.3. Machine and Deep Learning-based Analytical Modeling

In this section, we explore how various machine and deep learning techniques can be used for building analytical models for data-driven city services through intelligent decision-making utilizing relevant city data.

3.3.1. Machine Learning Modeling

Machine Learning (ML) is an important part of Artificial Intelligence (AI) that automates analytical model building, i.e., learning, exploring, and envisaging the smart city outcomes utilizing the collected city-data. Typically, machine learning models in the domain of smart cities can analyze the data, extract knowledge or useful insights, and eventually build the data-driven model to make intelligent decisions in various situations for the citizens. Most importantly, the data-driven decision-making ability of smart cities based on machine learning methods and corresponding systems can make the city services more accessible, efficient, and intelligent, and eventually improve the quality of the citizens' life. In the area of the machine, learning [11], various techniques are popular to analyze the data depending on the data characteristics and target outcome. For example -

- When particular targets are stated to be achieved from a certain set of inputs, supervised learning, i.e., "task-driven approach", is used. For instance, predicting energy usage [low, normal, high] in a certain region of the city from the historical usage data, classification techniques [11] can be used. Several common classification algorithms exist in the area of machine learning, such as Navies Bayes [46], Decision Trees [47] [48] [49], K-nearest neighbors [50], Support vector machines [51], Adaptive boosting [52], Logistic regression [53], and many more summarized in our earlier paper Sarker et al. [11]. Thus, the discovered insights based on these machine learning techniques can play a key role to perform classification tasks or predicting the future outcome utilizing the city data as well as detecting cyber-threats [54], such as predicting traffic jams under different traffic conditions, disease diagnosis, and prediction, detecting malware attacks or anomalies in the context of smart cities. Several feature engineering tasks, such as selecting the optimal number of features according to their impact on the outcome, generating principle components or new brand features from city data, or the tasks of context pre-modeling [6] can make the analytical city model more effective and actionable.

- Unsupervised learning [45], on the other hand, typically investigates the similarity in city data that can play a role to generate suggestions or recommendations for the citizens. For example, clustering algorithms such as K-means [55], K-medoids [56], Single linkage [57], Complete linkage [58], BOTS [59] can be used by taking into account certain similarity measures in citizens' preferences or their behavioral activities or usage. The clustering techniques can also be used to find anomalies or cyber-threats in the smart city data. In addition to these methods, association rules in the area of machine learning can also be used to build rule-based intelligent systems for city services. Several approaches such as AIS [60], Apriori [61], FP-Tree [62], RARM [63], Eclat [64] etc. can be used for building such smart city model for the citizens. Considering various types of contextual information such as temporal, spatial, social, etc. of the citizens and building a rule-based model can make the smart city model adaptive and more effective. Our earlier approach ABC-RuleMiner Sarker et al. [16] can play a key role to build such a context-aware smart city model by taking into account the current contexts of the citizens.

Furthermore, semi-supervised learning can be defined as a hybridization of the supervised and unsupervised techniques stated above, and it could be effective for improving the performance of city models when data is needed to be labeled automatically without human intervention. Reinforcement techniques are a sort of machine learning that distinguishes an agent by allowing it to create its own learning experiences by interacting with the environment directly, i.e., an environment-driven approach [11]. Thus, we can conclude that machine learning techniques are the key to building analytical model building depending on the data characteristics and target outcome, within the area of data-driven smart cities.

3.3.2. Deep Learning Modeling

Deep learning (DL) also refers to data-driven learning approaches that employ multi-layer neural networks and processing [12], and could be useful to handle large datasets in the environment of smart cities. For instance, to predict future values of air quality in a smart city a Long short-term memory (LSTM) network-based deep learning model is used in [65]. Similarly, several other discriminative deep learning techniques such as Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN or ConvNet), Recurrent Neural Networks (RNN), and their derivatives that are originated from an Artificial Neural Network (ANN) can be used in the learning process. For instance, in the smart city applications of transport flow, human mobility, parking, air quality, healthcare, etc. these deep learning models are used summarized in [66]. In addition, generative modeling techniques such as Generative Adversarial Network (GAN), Autoencoder (AE), Restricted Boltzmann Machine (RBM), Self-Organizing Map (SOM), and Deep Belief Network (DBN), as well as their derivatives [12], can be utilized as a preprocessing step for such supervised learning tasks, ensuring the model accuracy. By taking into account discriminative and generative tasks, DL techniques can be classified into three major categories, discussed briefly in our earlier paper Sarker et al. [12] with their functionalities. These are - (i) deep networks for unsupervised or generative learning, (ii) deep networks for supervised or discriminative learning, and (iii) deep networks for hybrid learning that could be useful in the context of smart cities depending on the data characteristics. For instance, a hybrid CNN-LSTM model has been employed for short-term individual household load forecasting, presented in [67].

Overall, machine learning and deep learning analytical models can play a crucial part in comprehending and interpreting actual occurrences with city data, depending on the intended smart city application and relevant data properties.

Therefore, we can conclude that learning models can improve the effectiveness of smart city platforms through intelligent city services, because of their learning capability from city data collected from relevant sources in the context of smart cities.

4. Smart City Services and Data-Driven Decision-Making

This section explores data-driven smart city services and systems in a variety of possible application domains in the context of smart cities, emphasizing the importance of data-driven decision-making within the scope of our study. Thus we have divided them into ten categories, such as smart environments, transportation, energy management, public safety, cybersecurity management, and so on:

(i) *Smart People or Citizens:* In smart cities, people are the primary users of smart devices and services. Smart city residents should interact and communicate with each other in order to share common and crucial online social experiences as well as physical space. Smart people should not only use services to communicate with others, but they should also provide data to them, where data science modeling can play a key role to analyze the historical data to make intelligent decisions through extracting insights or useful knowledge. For example, by integrating automatic sensor readings from cellphones with manual input from users, a crowdsourced weather application, can assess data on present and upcoming weather events. In addition, social networking platforms could be one of the most important sources of information on smart city systems and services, as well as public opinion and sentiment.

(ii) *Smart Governance:* In addition to smart people, mentioned above, smart governance might be another key topic in the context of a country's smart cities. It's about the future of public services and systems, which will include increased efficiency, community leadership, mobile working, and

 doi:10.20944/preprints202204.0044.v1

continuous development through innovation and technology, where data-driven decision-making can intelligently provide services based on the preferences of citizens. Through the use of social media or data from relevant organizations, data science modeling can assist governments in understanding current trends and demands of citizens in order to develop policies and, ultimately, establish a smart governmental administration. The parameters by which government actions are evaluated include efficiency, effectiveness, transparency, and collaboration. Overall, smart governance is the practice of employing technologies to create a collaborative, transparent, participative, communication-based, and sustainable environment for citizens and governments, where data-driven decision-making could be more effective rather than the traditional approach.

(iii) *Smart Environment and Resource Management:* Smart environments have recently gained prominence with the support of academics, and they are one of the most significant parts of smart cities. Researchers look at current environmental models and systems to improve monitoring, understanding, prediction, and management of city settings, with a focus on air quality, green and aquatic areas, pollution monitoring, waste management, energy efficiency, and so on, in order to create a smart environment. As IoT and ubiquitous sensing become more popular, researchers are aiming to develop cost-efficient technologies for acquiring fine-grained environmental data in real-time and their effective analysis for smart environments [40]. For example, water consumption analysis can help with the design of residential water systems as well as control systems that allow governments to manage water resources; similar management systems can be used with electricity and gas. Flood analysis can predict specified rain thresholds, alerting us to the possibility of flooding. Daily pollution analysis can be used to notify residents of smart cities about the level of pollution and whether certain activities should be disallowed as a result of it. Longer spans of time analysis may be useful in urban planning and traffic management. By analyzing meaningful monitoring data and understanding the patterns and correlations between environmental components, more accurate predictions can be made, as well as intelligent decision-making.

(iv) *Smart Homes, Buildings and Living:* Smart homes and buildings are considered one of the important application areas in smart cities. The reason is that most of these are related to various sectors of citizens' life ranging from their living environment at home to commercial purposes. In recent days, smart buildings can be conceptualized not only at the physical level but also at the virtual level. For instance, on a physical level, the smart buildings may have networking capabilities, power grid, transportation system, etc., while on a virtual level, it is related to information sharing, collaboration, or other utilities in a community. Moreover, managing security systems, surveillance, lighting control, automated operations, intelligent services, etc. are also taken into account as the components of smart homes and buildings. Thus, to enhance the quality of services, the utilization of knowledge discovered through using data science modeling can be used, to offer high-quality services to citizens. For example, for the betterment of people, a context-aware smart home regulates home appliances, energy consumption, lighting control, etc [68] [69]. Smart warehouses will likewise improve the efficiency of supply chain management, which also benefits community members [70]. Overall, the term 'smart living' refers to smart buildings for education, tourism, healthcare, and public safety, all of which can improve citizens' quality of life, and data-driven decision-making through extracting insights from the relevant sources is the key.

(v) *Smart Education*: Smart education services and applications are often flexible and intelligent for citizens' lifelong learning since they are capable of engaging individuals in active learning environments to adapt to society's and the environment's rapid changes. According to Tikhomirov

 doi:10.20944/preprints202204.0044.v1

et al., [71], the main dimensions of smart education are educational outcomes, ICT, and organizational dimensions. Thus, collecting data in the field such as related to the people (e.g., teachers, students, parents, administrators, etc.), infrastructures (e.g., institutes, computing facilities, etc.), and information (e.g. courses assessments, outcome, economic surveys, reports, etc.) can be a useful resource for smart education [26]. The ability to extract meaningful insights from data through proper processing will have a favorable impact on knowledge levels and teaching-learning tools for delivering or acquiring knowledge. Thus making better use of information and assessment could be a better way to improve the efficiency, effectiveness, and productivity of educational processes. Using data science modeling expertise in the context of smart cities can assist in the creation of a knowledge-based society, which will improve the nation's competitiveness.

(vi) *Smart Transportation, Parking and Traffic Lights:* Smart transportation systems promote traffic flow and residents' daily lives while also presenting public safety and environmental concerns. In order to develop a smart city, cutting-edge transportation systems are utilized to monitor a city's mobility, optimize traffic performance, and deliver superior transportation services. This can be accomplished by combining heterogeneous traffic data, mining traffic trajectories for traffic patterns, and evaluating and predicting traffic flow for intelligent decision-making. For this, it includes information systems based on data science modeling that first collect transport data about the traffic, vehicles, ticketing, and usage of different transport modes, as well as the sentiment or opinion of the citizens regarding the relevant services. For instance, a smart transportation system has been designed in Jan et al. [72] based on IoT and big data approaches. By providing precise location and routing information [73], the concept of IoT will boost smart transportation in smart cities. It will also allow people to make better decisions on their schedules, such as the

 doi:10.20944/preprints202204.0044.v1

degree of street congestion, alternate routes, transport medium, which can reduce their wasted transportation time [30]. Modern systems typically offer intelligent road networks, subway and metro train networks, public transportation, cycle routes, pedestrian paths, etc. [74].

Smart parking facilities are also considered another important part of traffic management systems as well as smart cities, due to the increasing number of vehicles, and the mismanagement of the parking space [75]. Thus, smart parking can provide real-time information according to the needs that can help to effectively find the parking slots, considering cost and time. For instance, a smart parking system results in high parking space utilization and fast parking spot finding time through the simulation results has been presented in [76]. Traffic flow-based IoT sensor data (e.g. vehicle speeds, density, waiting time, traffic jams, etc.) could be a valuable resource for smart traffic lights that can provide more data on traffic patterns [26]. For example, smart traffic lights and signals have reduced pollution by more than 20 percent, shown in [77]. Thus, using the knowledge of data science modeling in the context of smart cities can help to build such systems, which, can lead to better traffic flow under different traffic conditions.

(vii) *Smart Healthcare:* Smart healthcare is one of the most important parts of human life as well as in smart cities. Smart health typically combines smart technologies with smart IoT devices like smartphones or sensors. The devices collect data related to health issues using embedded sensors to extract valuable insights for the early detection of health issues. Smart health technology can also assist doctors, researchers, and health care professionals with better-personalized diagnoses and solutions by analyzing health-related data. The IoT devices are not only just monitor one's health issues but can also provide solutions by analyzing the collected data using data science modeling, which eventually, leads to a smarter way of providing healthcare services to improve the citizens' quality of life

[78]. For instance, Souri et al. [79] present an IoT-based student healthcare monitoring model to detect physiological and behavioral changes via smart healthcare technologies using machine learning methods.

(viii) *Smart Grid and Energy:* The majority of city energy systems are reliant on national energy systems. We can focus on various aspects of smart cities through data science modelings, such as forecasting energy demands suited to different locations; analyzing and demand forecasting, consumer behavior analysis, usage patterns at the individual and population levels; energy modeling for buildings, cities, and industry; energy generation optimization. Improving the efficiency of energy grids is another aspect of energy systems that are revolutionizing smart cities. It analyzes the behaviors of the suppliers as well as the consumers and thus could be an effective solution through data-driven decision-making in the smart grid system [80], while the traditional system is manually evaluated [81]. For this, smart sensors and meters may collect data on output, transmission, distribution networks, and customer access points to obtain granular data on current power production, use and faults [26]. Smart grids are being used to make energy distribution and transmission more efficient, reliable, and secure, as well as to satisfy increasing energy demands and minimize pollution. Thus, using the knowledge of data science modeling in the context of smart cities, including machine learning methods can help to build such systems, which can lead to improving the overall performance of the electric power system under different conditions.

(ix) *Emergency and Public Safety:* One of the most crucial areas of smart city research is emergency and public safety. For a long time, cities have been looking for solutions to deal with emergencies and risky situations caused by natural disasters (e.g., earthquakes, hurricanes), city facility malfunctions (e.g., power outages, bridge collapses), and human factors (e.g., automobile accidents, building fires). Emergency and public safety systems are growing smarter as the IoT becomes more widespread. The

IBM Intelligent Operations Center [82], for example, is a system for managing emergency occurrences such as natural and human-caused disasters. One significant source to monitor and use to govern city safety is public safety data, such as crime and accident statistical data. The police departments of several cities release criminal data. Citizens can use public safety data to get timely alerts, measure neighborhood safety, disseminate safety measurements and devices, and construct smart safety services to prevent asset and human loss. More timely and emergency safety data is expected to become publicly available in the near future, due to the advancement of smart real-time platforms. Moreover, information fusion from historical and real-time data could be more useful while making data-driven decisions in the context of emergency and public safety.

(x) *Cybersecurity Management:* Cyber-security concerns such as information leakage and malicious cyber-attacks in this field affect smart city behavior since numerous components of smart cities rely on information and communication technologies [83]. Thus, in a smart city, a single susceptible action by an individual or group might put the entire city at risk. Therefore, cyber security is needed to take into account the widespread acceptance of global smart city technology. Several cybersecurity issues include illegal access to information, anomalies or attacks such as malware, ransomware, botnet, phishing, social engineering, zero-day attacks, denial-of-service (DoS), distributed denial-of-service (DDoS), etc. [5] are considered serious concerns about dangers in the growth of smart environments, and may cause disruptions in such smart services. Thus, the smart city security analysis is important which is typically the process of detecting vulnerabilities, cybersecurity threats or attacks, and security concerns associated with smart city infrastructure assets, as well as countermeasures, to mitigate these threats. Several well-known security approaches such as access control [84], firewall [85], anti-malware [86], Sandbox [87], security information and event management (SIEM) [88], or cryptography

[89], are typically used to mitigate the issues. However, data science modeling is demanding security intelligence as well through its capability of extracting more actionable insights from the data to intelligently manage these issues.

Overall, we can conclude that as cities grow more connected, enormous economic, environmental, and quality-of-life benefits arise. However, with cyber everywhere, this connectivity brings with it its own set of privacy and security concerns. Therefore, the concept of data-driven smart cities could be a realistic option for delivering smart services and security in various application domains discussed above, where data-driven decision-making through data science modeling is the key.

5. Open Research Issues

Even though the smart city concept has received widespread acceptance and is being implemented in the real world, addressing current challenges in specific areas has become crucial to accomplish further progress. In the following, we highlight the open research issues identified within the scope of our study to broaden the knowledge base and provide directions for further research.

- In the real world, a smart city is not static; it's always changing and improving. As a result, smart cities have a high level of uncertainty because services operating in open, highly dynamic surroundings, as well as sensors, may fail. Furthermore, human behavior is unpredictable, which has an impact on city services. Thus the real-world collected data may have missing, inconsistency, ambiguous, anomalies, noise, or incorrect values [41]. In addition, several disruptive events such as gas leaks, bad weather, accidents, and many others may occur in the city area, which causes data uncertainty as well. As uncertainty increases in smart cities, data-driven modeling and decision-making grow more complex, and consequently, the quality of services as a whole would be degraded. Therefore effectively

 doi:10.20944/preprints202204.0044.v1

handling such uncertainty as well as sufficient data availability to ensure the quality of the data could be one of the primary issues.

- In the context of smart cities, most of the collected data is unstructured and untagged, collected from various sources, e.g., sensors or IoT devices [5]. Although proper tagging and structuring can improve the quality of the city data, data annotation tasks, e.g., categorization, tagging, or labeling of a large amount of raw data is a challenging and time-consuming tasks. Thus data-driven modeling for extracting generative features, identifying meaningful trends and structures, groupings in results become interesting to extract insights or useful knowledge from data. Several data-driven techniques such as clustering, density estimation, feature learning, anomaly detection, dimensionality reduction, discovering association rules, etc. in the area of machine learning [11] could be useful. Therefore, extracting actionable insights from the collected city data, and eventually designing an effective data-driven model could be a major issue in this domain.

- Semantic interoperability could be another direction of research in the context of smart cities. It is defined as the exchange of data with meaningful and understandable meanings that incorporates semantics in data by integrating self-described information packages. Data technologies are required for most data-related tasks, such as processing, analyzing, and extracting insights [10], whereas semantic technologies that extract and represent knowledge are required for meaning-related tasks, such as event detection, reasoning, and decision support in a specific city problem. Multiple data sources are needed to be merged to realize the potential of a smart city, where the data are heterogeneous and eventually, difficult to understand the meaning of the raw data. Thus to overcome the interoperability issues, a semantic data model, i.e., a knowledge-based model, is essential, which can associate relationships and discover new ones when needed. Therefore integration with semantic technologies and designing an effective knowledge-based model to support the relevant smart city application development is another issue in the area.

- Many time-sensitive applications such as smart and connected vehicles, smartphone applications, and many more in a smart city context require real-time or near-real-time data analytics. New analytic frameworks that allow advanced data analytics, as well as streaming data analytics, are required for these applications. Therefore, designing an effective data-driven approach to adapt the time-series model [59] for next-generation mobile, IoT, or resource-constrained devices and applications could be another major issue in the area.
- In some circumstances, the typical machine learning [11] and deep learning [12] techniques may not be effective to build an analytical city model. It depends on data characteristics, problem nature, as well as target solution. In a rule-based system, for example, the association rule learning technique [61] extracts redundant generation from the data, making the decision-making process complicated and unproductive [16]. Therefore, a deeper understanding of the strengths and limitations of existing learning methods is required, and proposing new techniques as well as their ensembles could be a promising direction for data-driven smart city research.
- Due to security and privacy issues, data may not be transported to the cloud. To take advantage of real-time intelligence, smart city applications may require lightweight machine learning algorithms that may be deployed on resource-constrained devices. Even though the traditional machine learning methods [11] are capable of analyzing vast amounts of data, lightweight modeling is necessary for resource-constrained devices due to their high computational cost and considerable memory overhead. Therefore another key issue in the context of smart cities could be the development of lightweight learning approaches to adapt the data-driven model for next-generation mobile, IoT, or resource-constrained devices and applications.

 doi:10.20944/preprints202204.0044.v1

- Another significant issue is that the environment of smart city applications is not static may change over time. In the real-world scenario, recent pattern-based modeling and analysis is more likely to be interesting and significant than older ones for predicting the future, highlighted in our earlier paper Sarker et al. [90]. Thus the concept of recency mining, i.e., city modeling based on extracted recent patterns from data could be more effective. Therefore the issue is how we can incorporate the recency in smart city services as well as incremental learning, i.e., to support continuous learning and updating.

- In a smart city, a single individual's or organization's vulnerable action might put the entire city at risk. Cyber-security concerns such as information leakage and harmful cyber-attacks [5] in this field affect smart city behavior since numerous components of smart cities rely on information and communication technologies. Data-driven machine learning modeling could be one effective solution to detect anomalies or cyber-attacks that have been shown through experimental analysis in our earlier paper Sarker et al. [54]. However, in smart cities, security issues are mostly application-oriented. Therefore more advanced and data analytics-based innovative techniques according to the application perspective are needed to ensure the cyber safety and security of smart city applications, which could be another significant issue in the context of smart cities.

- Integrating context information such as temporal, spatial, environmental, or social context, etc. with raw data is crucial for extracting more value from the data and performing faster and more precise reasoning and actuation, highlighted in our earlier book Sarker et al. [14]. Such contextual information may influence users' behavioral activities in real-world scenarios which may vary from user to user. Thus, the citizens must be provided with the appropriate contextual information about the characteristics and processes of their urban surroundings in various contexts of smart cities, such as socio-economic activities, quality of life, and citizen well-being.

doi:10.20944/preprints202204.0044.v1

Therefore, effectively and efficiently incorporating context-awareness in smart city services could be a significant issue and research direction.

- Designing a data-driven analytical framework that allows for data science modeling, as explained in Section 3 for Smart Cities, is the most crucial task for a data-driven smart city solution. Analytical frameworks are designed to help analysts organize their ideas and concepts logically and methodically, where the models support and facilitate the process of making sense of things. Thus advanced analytical city modeling based on machine learning or deep learning discussed earlier, might be included in such a system to allow it to handle the target issue. Furthermore, to construct a context-aware dynamic adaptive framework, context-aware machine learning [14] can be integrated. Therefore a well-designed context-aware smart city framework, as well as experimental assessment, is a very important direction and a big challenge as well.

In summary, this article has identified the above ten potential future directions in the field of data-driven smart cities. This can also help researchers conduct a deeper exploration of the application's hidden and unexpected issues, resulting in more accurate and realistic outcomes. In conclusion, addressing the aforementioned concerns and contributing to the development of effective and efficient approaches may result in more automated and efficient applications in the context of smart cities.

6. Conclusion

We have explored how smart city data science contributes to data-driven intelligent decision-making in smart city systems and services in this paper, which was motivated by the growing significance of data science in today's computing. We concentrated on extracting insights from city data, starting with the research design and moving on to suggestions for data-driven smart city solutions. Thus, we have also explored how different machine learning

 doi:10.20944/preprints202204.0044.v1

and deep learning techniques can be employed to develop analytical models in data-driven smart cities to fulfill the needs of the people. The concept of data-driven smart cities encompasses a wide range of real-world services aimed at improving people's lives, including smart environments, education, healthcare, transportation, energy management, public safety, cybersecurity, and so on, discussed briefly in this paper. Therefore, a data-driven smart city can secure the availability of city resources in terms of social, economic, and environmental aspects, thereby improving citizens' quality of life.

Overall, we aimed to provide an insight of smart city data science conceptualization, thinking, modeling, and processing, where the applicability of machine learning approaches to data-driven intelligent decision making in smart city services has also been explored. Finally, the challenges encountered as well as open research issues within the scope of our study have been emphasized, allowing researchers and academics to pursue further research in the areas specified. We believe that our study on data-driven smart cities points down a promising path that could be used as a reference guide for academics and researchers, as well as smart city designers at application level, leading to increased involvement in academics, industries, and governments of a country.

References

[1] B. N. Silva, M. Khan, K. Han, Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities, Sustainable Cities and Society 38 (2018) 697–713.

[2] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, B. Sikdar, A survey on iot security: application areas, security threats, and solution architectures, IEEE Access 7 (2019) 82721–82743.

[3] E. O'Dwyer, I. Pan, S. Acha, N. Shah, Smart energy systems for sustainable smart cities: Current developments, trends and future directions, Applied energy 237 (2019) 581–597.

[4] Z. Ullah, F. Al-Turjman, L. Mostarda, R. Gagliardi, Applications of artificial intelligence and machine learning in smart cities, Computer Communications (2020).

[5] I. H. Sarker, A. I. Khan, Y. B. Abushark, F. Alsolami, Internet of things (iot) security intelligence: A comprehensive overview, machine learning solutions and research directions, Mobile Networks and Applications (2022) 1–17.

[6] I. H. Sarker, H. Alqahtani, F. Alsolami, A. I. Khan, Y. B. Abushark, M. K. Siddiqui, Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive modeling, Journal of Big Data 7 (1) (2020) 1–23.

[7] B. Ślusarczyk, Industry 4.0: Are we ready?, Polish Journal of Management Studies 17 (2018).

[8] C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, P. Williams, Foundations for smarter cities, IBM Journal of research and development 54 (4) (2010) 1–16.

[9] S. Kondepudi, V. Ramanarayanan, A. Jain, G. Singh, N. Nitin Agarwal, R. Kumar, R. Singh, P. Bergmark, T. Hashitani, P. Gemma, et al., Smart sustainable cities analysis of definitions, The ITU-T focus group for smart sustainable cities (2014).

[10] I. H. Sarker, Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective, SN Computer Science (2021).

[11] I. H. Sarker, Machine learning: Algorithms, real-world applications and research directions, SN Computer Science (2021).

[12] I. H. Sarker, Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions, SN Computer Science 2 (6) (2021) 1–20.

[13] I. H. Sarker, Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems, SN Computer Science (2022).

[14] I. H. Sarker, A. Colman, J. Han, P. Watters, Context-Aware Machine Learning and Mobile Data Analytics: Automated Rule-based Services with Intelligent Decision-Making, Springer, 2022.

[15] I. H. Sarker, Context-aware rule learning from smartphone data: survey, challenges and future directions, Journal of Big Data 6 (1) (2019) 1–25.

[16] I. H. Sarker, A. Kayes, Abc-ruleminer: User behavioral rule-based machine learning method for context-aware intelligent services, Journal of Network and Computer Applications 168 (2020) 102762.

[17] I. H. Sarker, A. Kayes, S. Badsha, H. Alqahtani, P. Watters, A. Ng, Cybersecurity data science: an overview from machine learning perspective, Journal of Big Data 7 (1) (2020) 1–29.

[18] A. Caragliu, C. Del Bo, P. Nijkamp, Smart cities in europe, Journal of urban technology 18 (2) (2011) 65–82.

[19] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, H. J. Scholl, Understanding smart cities: An integrative framework, in: 2012 45th Hawaii international conference on system sciences, IEEE, 2012, pp. 2289–2297.

[20] V. Albino, U. Berardi, R. M. Dangelico, Smart cities: Definitions, dimensions, performance, and initiatives, Journal of urban technology 22 (1) (2015) 3–21.

[21] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, Y. Portugali, Smart cities of the future, The European Physical Journal Special Topics 214 (1) (2012) 481–518.

[22] H. Ahvenniemi, A. Huovila, I. Pinto-Seppä, M. Airaksinen, What are the differences between sustainable and smart cities?, Cities 60 (2017) 234–245.

[23] A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, Internet of things for smart cities, IEEE Internet of Things journal 1 (1) (2014) 22–32.

[24] E. Ismagilova, L. Hughes, Y. K. Dwivedi, K. R. Raman, Smart cities: Advances in research—an information systems perspective, International Journal of Information Management 47 (2019) 88–100.

[25] M. Batty, Big data, smart cities and city planning, Dialogues in human geography 3 (3) (2013) 274–279.

[26] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, J. Al-Jaroodi, Applications of big data to smart cities, Journal of Internet Services and Applications 6 (1) (2015) 25.

[27] G. P. Hancke, G. P. Hancke Jr, et al., The role of advanced sensing in smart cities, Sensors 13 (1) (2013) 393–425.

[28] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, A. Oliveira, Smart cities and the future internet: Towards cooperation frameworks for open innovation, in: The future internet assembly, Springer, Berlin, Heidelberg, 2011, pp. 431–446.

[29] I. U. Din, M. Guizani, J. J. Rodrigues, S. Hassan, V. V. Korotaev, Machine learning in the internet of things: Designed techniques for smart cities, Future Generation Computer Systems 100 (2019) 826–843.

[30] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, A. P. Sheth, Machine learning for internet of things data analysis: A survey, Digital Communications and Networks 4 (3) (2018) 161–175.

[31] J. T. d. Souza, A. C. d. Francisco, C. M. Piekarski, G. F. d. Prado, Data mining and machine learning to promote smart cities: A systematic review from 2000 to 2018, Sustainability 11 (4) (2019) 1077.

[32] T. Yigitcanlar, N. Kankanamge, M. Regona, A. Ruiz Maldonado, B. Rowan, A. Ryu, K. C. Desouza, J. M. Corchado, R. Mehmood, R. Y. M. Li, Artificial intelligence technologies and related urban planning and development concepts: How are they perceived and utilized in australia?, Journal of Open Innovation: Technology, Market, and Complexity 6 (4) (2020) 187.

[33] T. Yigitcanlar, L. Butler, E. Windle, K. C. Desouza, R. Mehmood, J. M. Corchado, Can building "artificially intelligent cities" safeguard humanity from natural disasters, pandemics, and other catastrophes? an urban scholar's perspective, Sensors 20 (10) (2020) 2988.

[34] City of new york. 2017. nyc open data. retrieved from: https://opendata.cityofnewyork.us/.

[35] T. Bakıcı, E. Almirall, J. Wareham, A smart city initiative: the case of barcelona, Journal of the knowledge economy 4 (2) (2013) 135–148.

[36] Beijing city lab. 2017. bcl: Beijing city lab. retrieved from: https://www.beijingcitylab.com/.

[37] Chicago. 2019. chicago data port. retrieved from https://data.cityofchicago.org/.

[38] data.gov.uk. 2017. opening up government in uk. retrieved from: https://data.gov.uk/.

[39] R. Kosala, E. Adi, Harvesting real time traffic information from twitter, Procedia Engineering 50 (2012) 1–11.

[40] M. Ma, S. M. Preum, M. Y. Ahmed, W. Tärneberg, A. Hendawi, J. A. Stankovic, Data sets, modeling, and decision making in smart cities: A survey, ACM Transactions on Cyber-Physical Systems 4 (2) (2019) 1–28.

[41] I. H. Sarker, A machine learning based robust prediction model for real-life mobile phone data, Internet of Things 5 (2019) 180–193.

[42] M. L. Waskom, Seaborn: statistical data visualization, Journal of Open Source Software 6 (60) (2021) 3021.

[43] X. Qin, Y. Luo, N. Tang, G. Li, Making data visualization more efficient and effective: a survey, The VLDB Journal 29 (1) (2020) 93–117.

[44] J. Liu, T. Tang, W. Wang, B. Xu, X. Kong, F. Xia, A survey of scholarly data visualization, Ieee Access 6 (2018) 19205–19221.

[45] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.

[46] G. H. John, P. Langley, Estimating continuous distributions in bayesian classifiers, in: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

[47] J. R. Quinlan, C4.5: Programs for machine learning, Machine Learning (1993).

[48] I. H. Sarker, A. Colman, J. Han, A. I. Khan, Y. B. Abushark, K. Salah, Behavdt: A behavioral decision tree learning to build user-centric context-aware predictive model, Mobile Networks and Applications (2019) 1–11.

[49] I. H. Sarker, Y. B. Abushark, F. Alsolami, A. I. Khan, Intrudtree: A machine learning based cyber security intrusion detection model, Symmetry 12 (5) (2020) 754.

[50] D. W. Aha, D. Kibler, M. K. Albert, Instance-based learning algorithms, Machine learning 6 (1) (1991) 37–66.

[51] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, K. R. K. Murthy, Improvements to platt's smo algorithm for svm classifier design, Neural computation 13 (3) (2001) 637–649.

[52] Y. Freund, R. E. Schapire, et al., Experiments with a new boosting algorithm, in: Icml, Vol. 96, Citeseer, 1996, pp. 148–156.

[53] S. Le Cessie, J. C. Van Houwelingen, Ridge estimators in logistic regression, Journal of the Royal Statistical Society: Series C (Applied Statistics) 41 (1) (1992) 191–201.

[54] I. H. Sarker, Cyberlearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks, Internet of Things 14 (2021) 100393.

[55] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, 1967.

[56] L. Rokach, A survey of clustering algorithms, in: Data Mining and Knowledge Discovery Handbook, Springer, 2010, pp. 269–298.

[57] P. H. Sneath, The application of computers to taxonomy, Journal of General Microbiology 17 (1) (1957).

[58] T. Sorensen, method of establishing groups of equal amplitude in plant sociology based on similarity of species, Biol. Skr. 5 (1948).

[59] I. H. Sarker, A. Colman, M. A. Kabir, J. Han, Individualized time-series segmentation for mining mobile phone user behavior, The Computer Journal 61 (3) (2018) 349–368.

[60] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, in: ACM SIGMOD Record, Vol. 22, ACM, 1993, pp. 207–216.

[61] R. Agrawal, R. Srikant, et al., Fast algorithms for mining association rules, in: Proc. 20th int. conf. very large data bases, VLDB, Vol. 1215, 1994, pp. 487–499.

[62] J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, in: ACM Sigmod Record, Vol. 29, ACM, 2000, pp. 1–12.

[63] A. Das, W.-K. Ng, Y.-K. Woon, Rapid association rule mining, in: Proceedings of the tenth international conference on Information and knowledge management, ACM, 2001, pp. 474–481.

[64] M. J. Zaki, Scalable algorithms for association mining, IEEE transactions on knowledge and data engineering 12 (3) (2000) 372–390.

[65] İ. Kök, M. U. Şimşek, S. Özdemir, A deep learning model for air quality prediction in smart cities, in: 2017 IEEE International Conference on Big Data (Big Data), IEEE, 2017, pp. 1983–1990.

[66] Q. Chen, W. Wang, F. Wu, S. De, R. Wang, B. Zhang, X. Huang, A survey on an emerging area: Deep learning for smart city data, IEEE Transactions on Emerging Topics in Computational Intelligence 3 (5) (2019) 392–410.

[67] M. Alhussein, K. Aurangzeb, S. I. Haider, Hybrid cnn-lstm model for short-term individual household load forecasting, IEEE Access 8 (2020) 180544–180557.

[68] M. Khan, B. N. Silva, C. Jung, K. Han, A context-aware smart home control system based on zigbee sensor network., Ksii Transactions on Internet & Information Systems 11 (2) (2017).

[69] M. Khan, B. N. Silva, K. Han, A web of things-based emerging sensor network architecture for smart control systems, Sensors 17 (2) (2017) 332.

[70] S. Jabbar, M. Khan, B. N. Silva, K. Han, A rest-based industrial web of things' framework for smart warehousing, The Journal of Supercomputing 74 (9) (2018) 4419–4433.

[71] V. Tikhomirov, N. Dneprovskaya, E. Yankovskaya, Three dimensions of smart education, in: Smart Education and Smart e-Learning, Springer, 2015, pp. 47–56.

[72] B. Jan, H. Farman, M. Khan, M. Talha, I. U. Din, Designing a smart transportation system: An internet of things and big data approach, IEEE Wireless Communications 26 (4) (2019) 73–79.

[73] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, M. Guizani, Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges, IEEE Wireless Communications 23 (5) (2016) 10–16.

[74] S. P. Mohanty, U. Choppali, E. Kougianos, Everything you wanted to know about smart cities: The internet of things is the backbone, IEEE Consumer Electronics Magazine 5 (3) (2016) 60–70.

[75] A. Khanna, R. Anand, Iot based smart parking system, in: 2016 International Conference on Internet of Things and Applications (IOTA), IEEE, 2016, pp. 266–270.

[76] G. Yan, W. Yang, D. B. Rawat, S. Olariu, Smartparking: A secure and intelligent parking system, IEEE intelligent transportation systems magazine 3 (1) (2011) 18–30.

[77] J. L. Galán-García, G. Aguilera-Venegas, P. Rodríguez-Cielos, An accelerated-time simulation for traffic flow in a smart city, Journal of Computational and Applied Mathematics 270 (2014) 557–563.

[78] M. Hassanalieragh, A. Page, T. Soyata, G. Sharma, M. Aktas, G. Mateos, B. Kantarci, S. Andreescu, Health monitoring and management using internet-of-things (iot) sensing with cloud-based processing: Opportunities and challenges, in: 2015 IEEE International Conference on Services Computing, IEEE, 2015, pp. 285–292.

[79] A. Souri, M. Y. Ghafour, A. M. Ahmed, F. Safara, A. Yamini, M. Hoseyninezhad, A new machine learning-based healthcare monitoring model for student's condition diagnosis in internet of things environment, SOFT COMPUTING (2020).

[80] R. Ak, O. Fink, E. Zio, Two machine learning approaches for short-term wind speed time-series prediction, IEEE transactions on neural networks and learning systems 27 (8) (2015) 1734–1747.

[81] Y. Zhang, T. Huang, E. F. Bompard, Big data analytics in smart grids: a review, Energy Informatics 1 (1) (2018) 8.

[82] A. Bhowmick, E. Francellino, L. Glehn, R. Loredo, P. Nesbitt, S. W. Yu, S. Mishr, Ibm intelligent operations center for smarter cities administration guide (2012).

[83] C. Ma, Smart city and cyber-security; technologies used, leading challenges and future recommendations, Energy Reports 7 (2021) 7999–8012.

[84] H. Qi, X. Di, J. Li, Formal definition and analysis of access control model based on role and attribute, Journal of information security and applications 43 (2018) 53–60.

[85] J. Yin, Firewall policy management, uS Patent 9,338,134 (May 10 2016).

[86] Y. Xue, G. Meng, Y. Liu, T. H. Tan, H. Chen, J. Sun, J. Zhang, Auditing anti-malware tools by evolving android malware and dynamic loading technique, IEEE Transactions on Information Forensics and Security 12 (7) (2017) 1529–1544.

[87] T. Hunt, Z. Zhu, Y. Xu, S. Peter, E. Witchel, Ryoan: A distributed sandbox for untrusted computation on secret data, ACM Transactions on Computer Systems (TOCS) 35 (4) (2018) 1–32.

[88] M. Irfan, H. Abbas, Y. Sun, A. Sajid, M. Pasha, A framework for cloud forensics evidence collection and analysis using security information and event management, Security and Communication Networks 9 (16) (2016) 3790–3807.

[89] O. G. Abood, S. K. Guirguis, A survey on cryptography algorithms, International Journal of Scientific and Research Publications 8 (7) (2018) 410–415.

[90] I. H. Sarker, A. Colman, J. Han, Recencyminer: mining recency-based personalized behavior from contextual smartphone data, Journal of Big Data 6 (1) (2019) 1–21.