Smart City Based on Open Data: A Survey

Karidja Dominique Christelle Adje, Asma Ben Letaifa, Majed Haddad, Oussama Habachi

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 May 2023, accepted 30 May 2023, date of publication 6 June 2023, date of current version 13 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3283436

SURVEY

Smart City Based on Open Data: A Survey

KARIDJA DOMINIQUE CHRISTELLE ADJE[1,3], ASMA BEN LETAIFA[1], MAJED HADDAD[2], AND OUSSAMA HABACHI[3]

[1]Mediatron Research Laboratory, SUP'COM, University of Carthage, Tunis 77-1054, Tunisia
[2]LIA, Avignon University, 84029 Avignon, France
[3]LIMOS, University of Clermont Auvergne, 63000 Clermont-Ferrand, France

Corresponding author: Karidja Dominique Christelle Adje (dominique.adje@supcom.tn)

This work was supported in part by the French National Research Agency under Grant ANR-20-CE25-000.

ABSTRACT Open data are gold mines because they can be used to create services that develop a smart city while improving users' living conditions. Several research works go in this direction, presenting open data impact in the smart city for some, while others have focused on data processing methods. We have therefore deemed it necessary to make a state of the art on these different issues. The particularity of our study is that it shows the link between open data and smart city in all its aspects, describing what kind of open data is suitable for the smart city, how it is important for its development, and how these open data are processed to create services. Thus, in this article, we first present a review of existing surveys since 2015. Then, we present different smart city dimensions based on open data as well as some applications, and we detail how to process these data. We end with a list of open data sources as well as some challenges and solutions related to smart city services.

INDEX TERMS Data mining, machine learning (ML), open data, smart city.

I. INTRODUCTION

A city is not only limited to physical urban spaces (e.g., places, and buildings) but also extends to systems, structures, networks, flows, and processes [1]. This is why Ahlers [2] defines a smart city as a livable, participative, and sustainable city. To develop such a city, open data produced by the different actors of the urban ecosystem play a crucial role [3], since their integration and valorization enable the development of high socio-economic impact services. The data are produced in a digital urban space that is composed of physical spaces (infrastructures, buildings, etc.), social spaces (government, population, organization, culture, etc.), and cyber spaces (internet data, communication data, etc.) [4]. The data are accessible through three main ways: official data portals (via the internet), big data initiatives (obtained explicitly or implicitly via crawling techniques), and the broader open data community [5]. However, not all open data are smart city-oriented. Prieto et al. [6] established 14 open data categories in the smart city to avoid ambiguity.

Innovative service development in a smart city first requires sharing these data categories, and then processing them efficiently using sophisticated data analytics techniques, including data mining and ML techniques. In this vein, in the literature, several studies have pointed out the importance and necessity of promoting data sharing and analysis using data mining and ML techniques. Our survey aims to combine all these aspects. So, our contribution is first, to provide an overview of existing surveys in the above-mentioned fields. Second, we introduce smart city dimensions as well as the related open data categories. In this section, we also present applications of the smart city. Third, all data analysis steps are detailed with an explanation of data mining and ML techniques used for each data type. Fourth, we provide a wide list of open data sources. Finally, some challenges and solutions related to smart city services are presented.

The remainder of this survey is organized as illustrated in Fig. 1. The next section presents corresponding existing surveys or reviews of open data and a smart city. Section III introduces smart city dimensions and applications based on open data analytics. Section IV outlines the data analytics concept, detailing the entire process with data mining and ML

The associate editor coordinating the review of this manuscript and approving it for publication was Sathish Kumar.

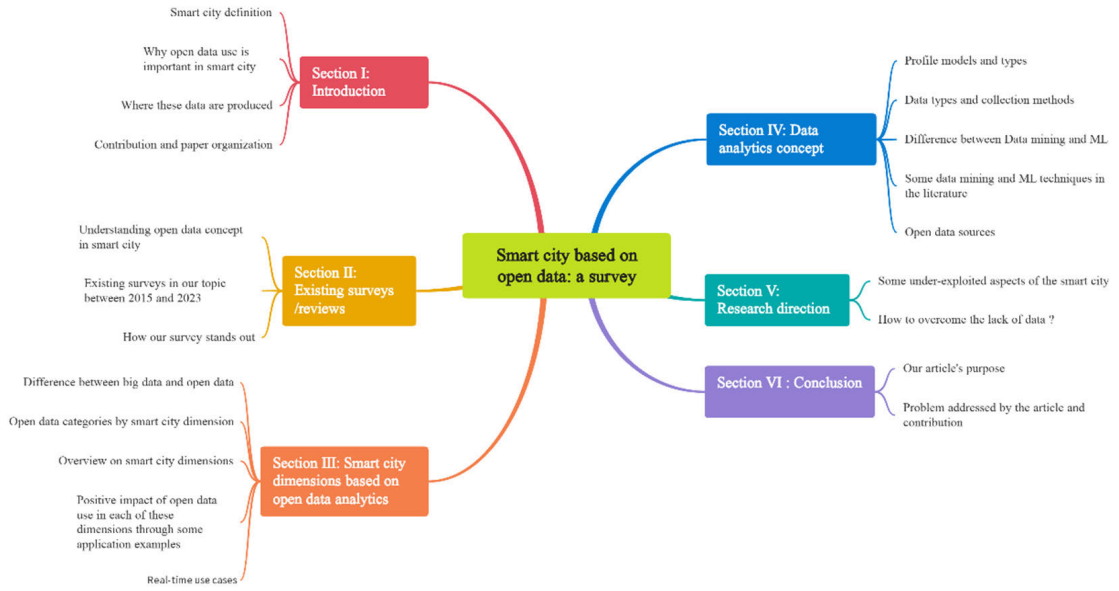

FIGURE 1. Survey taxonomy.

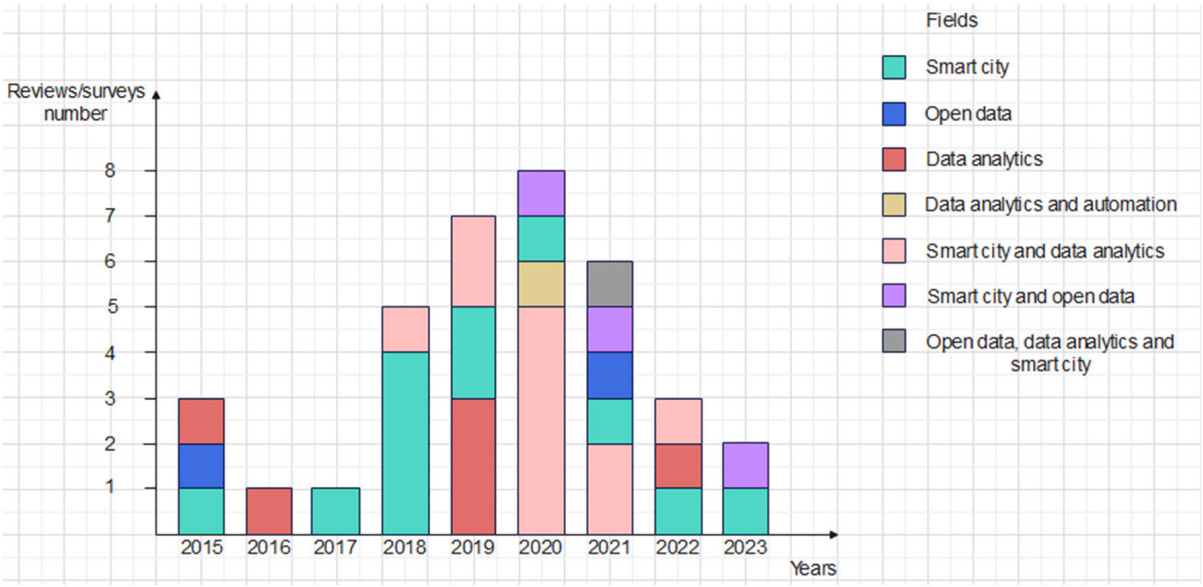

FIGURE 2. Statistics of existing surveys/reviews by our interest fields.

techniques. Section V presents future research directions in a smart city. We conclude with section VI.

II. EXISTING SURVEYS/REVIEWS

The smart city concept is in vogue more than ever. However, since this concept is rather blurred, several surveys in the literature have tried to define it clearly [7], [8], [9], [10]. The common idea of these papers is to present a smart city as a city characterized by sustainability, urbanization, smartness, and quality of life. On the other hand, data produced in a smart city are getting more attention from researchers as they are very valuable [11], [12], [13], [14]. Indeed, in a city, each habitant or actor produces data when performing an activity that represents his/her habits, interests, etc. [1]. These data must be accessible (open data) via digital services, to achieve several innovations that improve the city's functioning and its population's well-being [3]. In this regard, some reviews have shown the positive impact of using open data in a smart city [15], [16] without, however, clearly presenting techniques for exploiting these data. Other reviews have instead focused on data analytics techniques [17], [18], [19], [20], [21], [22]. In order to have an overview of existing surveys in this broad scope, we present in Fig. 2 a statistic of surveys or reviews by year and field, and Table 1 compares them to the fields covered in our paper.

For our survey, we gathered and synthesized articles that deal with at least one of our key themes: ''smart city'', ''open data'', ''data analysis steps'', ''profiling'', ''machine learning techniques'', and ''data mining techniques''.

As we can see in Table 1, no survey clearly presents what types of open data are suitable for the smart city. However, some of them [17], [20], [21], [22], [23], [29], [39] still cover most of our interest areas. We will therefore make a brief analysis of these surveys.

As we can see in Table 1 and Table 2, there is no survey that investigates ''smart city applications'', ''open data categories adapted to the smart city'', ''data analytics process'', ''data mining and ML techniques'', and ''open data sources'', all combined with an overview of existing surveys. Therefore, our survey meets this expectation by covering all these aspects.

III. SMART CITY DIMENSIONS BASED ON OPEN DATA ANALYTICS

A smart city is a sustainable city designed for socio-economic development and improvement of the quality of life, thanks to several means including ICTs (information and communication technologies) [43]. Thus, Giffinger and Gudrun [44], and Ma et al. [4] classified the smart city into six dimensions such as smart living, smart mobility, smart people, smart governance, smart environment, and smart economy. To develop these different fields, researchers use open data to identify the challenges and problems faced by populations, in order to create innovative services. Why specifically open data and not big data? Indeed, these 2 concepts are closely related, which can cause quite a lot of confusion. But there is a significant difference between them: open data are characterized by their accessibility and use while big data are characterized by their velocity, volume and variety and may not be publicly accessible. To be more precise, the open data concept makes big data more useful, more democratic, and less threatening [45]. However, not all open data are useful for the smart city, so Prieto et al. [6] broadly established 14 open data categories suitable for the smart city. We have therefore classified these 14 categories according to the six dimensions of the smart city in Fig. 3.

A. SMART MOBILITY

Smart mobility represents the strategies and techniques developed to manage and facilitate people's mobility.

1) TOURISM AND BEST ROUTES

In order to address mobility in the smart city, several works have focused on improving route planners to allow easy and fast travel, and tourism by proposing systems to generate POIs (point of interest) lists for the population. In this regard, Vázquez-Salceda et al. [46] established a system named Superhub that uses humans as sensors to collect useful information (location, movement flow, average trip times, interest, etc.) in order to not only get a better overview of the cities but also to deduce users' preferences through their profiling. Such information, coupled with weather data, allow their system to improve user mobility by generating an intelligent and personalized planner of opportunistic routes and POIs. In this same mobility recommendation perspective, Asad et al. [47] investigated passenger health safety in the railway system, due to covid 19 advent. The goal was to minimize its propagation by recommending appropriate itineraries and times to vulnerable travelers. In the tourism domain, Logesh et al. [48] considered that improvements could still be made in the recommendation systems. For example, to generate a POIs list for users, Logesh et al. [48] not only took into account their preferences and contextual information (e.g., weather) but also used their demographic information, and exploited the relationships between users in a group. The latter technique appears to be effective in expanding the POIs list since two persons with similar profiles are more likely to appreciate the same tourist destinations. Meanwhile, Abbasi-Moud et al. [49] proposed an approach limited to couple users' preferences with contextual information such as weather, time, and location for recommending POIs that are suitable for them.

2) ROAD TRAFFIC MANAGEMENT AND SAFETY

Beyond routes and POIs, good user mobility management also relies on shared transport optimization. For this purpose, Jäppinen et al. [50] implemented a bicycle-sharing system in Helsinki city using available population and trip data. Their results showed that combining a bicycle-sharing system with the existing traditional public transport means reduced travel time by an average of 10% in the entire region. Also in the urban traffic management context, Kong et al. [51] proposed a shared bus profiling scheme that is more suitable for users' requirements about route planning.

In order to ensure comfortable mobility and especially to prevent traffic accidents, Cooperative-Intelligent Transportation Systems have become key elements in transportation systems. These systems allow inter-vehicular and vehicular-to-environment communications to manage efficiently traffic, with the aim of ensuring road safety. As part of these systems, IoV (internet of vehicles) extends over a broader network involving entities such as humans, objects, and other heterogeneous networks [52]. Because execution time is really crucial in these systems, Chen et al. [53] proposed an offloading scheme to minimize a vehicular task's execution time. Although these systems are very important for traffic management, it should not be neglected that these traditional vehicles have a negative impact on the environment due to their dependence on fuel; e.g. in 2021, they caused an 8% increase in CO2 emissions compared with 2020 [54]. This is why smart mobility and smart environment work closely together. To develop smart mobility while respecting the smart environment requirements, solutions based on electric vehicles have been proposed [55], promoting sustainable and healthy lifestyles. Besides electric vehicles, Sanchez-Iborra et al. [56] proposed to integrate

TABLE 1. Comparative table of existing surveys/reviews according to our interest fields.

Surveys	Smart City	Open Data	Open Data categories for smart city	Data analytics process	Data mining techniques	ML techniques
[7]	Yes	-	-	-	-	-
[8]	Yes	-	-	Yes	-	-
[9]	Yes	Yes	-	-	-	-
[10]	Yes	-	-	-	-	-
[11]	Yes	-	-	Yes	Yes	-
[12]	Yes	Yes	-	-	-	-
[13]	-	Yes	-	-	-	-
[14]	-	Yes	-	-	-	-
[15]	Yes	Yes	-	-	-	-
[16]	Yes	Yes	-	-	-	-
[17]	Yes	-	-	-	Yes	Yes
[18]	Yes	-	-	-	-	Yes
[19]	Yes	Yes	-	-	-	Yes
[20]	Yes	Yes	-	Yes	Yes	Yes
[21]	Yes	Yes	-	-	-	Yes
[22]	Yes	-	-	Yes	-	Yes
[23]	Yes	Yes	-	Yes	Yes	-
[24]	-	-	-	Yes	Yes	-
[25]	Yes	-	-	-	-	-
[26]	Yes	-	-	-	-	-
[27]	Yes	-	-	Yes	-	-
[28]	Yes	-	-	-	-	-
[29]	-	-	-	Yes	Yes	Yes
[30]	-	-	-	Yes	Yes	Yes
[31]	-	-	-	Yes	Yes	Yes
[32]	Yes	-	-	-	-	-
[33]	Yes	-	-	-	Yes	Yes
[34]	-	-	-	-	Yes	-
[35]	Yes	Yes	-	-	-	Yes
[36]	Yes	-	-	-	Yes	Yes
[37]	Yes	-	-	-	Yes	Yes
[38]	Yes	-	-	-	-	Yes
[39]	Yes	Yes	-	-	Yes	Yes
[40]	-	-	-	Yes	Yes	Yes
[41]	Yes	-	-	-	-	-
[42]	Yes	Yes	-	-	-	-

two-wheeled eco-friendly personal vehicles (e.g., bicycles, motorcycles, segways, etc.) into these existing C-ISTs. Their system used an OBU (on-board unit), the cloud, LoRaWan, and NB-IoT (narrow band-internet of things) communication technologies. Their solution is called eco-efficient mobility and pays particular attention to the road safety of these two-wheeled vehicles, as they are more vulnerable to traffic accidents than traditional vehicles.

TABLE 2. Surveys closer to ours.

Surveys/Reviews	Summaries	Differences
2015 [23]	This article shows the importance of the Linked Open Data (LOD) mining approach in the processing of complex and heterogeneous data.	It is more focused on open-source tools for non-experts in data mining and programming, it does not address ML techniques. Open data categories adapted to the smart city are not specified and it does not present clear statistics on existing surveys in its scope. No open data sources are listed.
2018 [20]	This article presents a state-of-the-art of smart city-oriented data science. The authors present data production and analysis taxonomies as well as the resulting services.	The open data aspect is briefly mentioned. Open data categories adapted to the smart city are not specified and it does not present clear statistics on existing surveys in its scope.
2019 [17]	This article presents data mining and ML techniques used between 2000 and 2018.	Lacks a clear organization of these techniques. Focused only on smart mobility and smart environment works without addressing other dimensions of the smart city. Open data categories adapted to the smart city are not specified and this article does not present clear statistics on existing surveys in its scope. No open data sources are listed.
2019 [29]	This review presents clear data analysis taxonomy for establishing user profiles.	The authors make neither a link with the smart city nor with open data. Open data categories adapted to smart city are not specified and they do not present clear statistics on existing surveys in their scope.
2021 [21]	This survey focuses on smart city applications based on open data and ML.	It does not focus on data mining techniques in a holistic way, it does not detail the data analytics process. Open data categories adapted to the smart city are not specified. No open data sources are listed.
2022 [22]	The authors showed how data science contributes to developing smart city applications through ML.	This paper explains data analysis steps without specifying associated techniques. Open data categories adapted to the smart city are not specified and it does not present clear statistics on existing surveys in its scope. No open data sources are listed.
2022 [39]	This paper conducts a comprehensive study on the use of Graph Neural Networks for traffic forecasting	It is focused only on smart mobility. Open data categories adapted to the smart city are not specified and it does not present clear statistics on existing surveys in its scope.

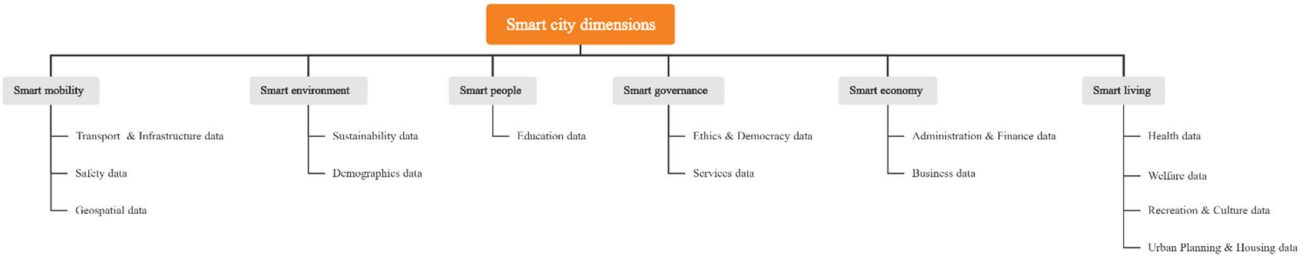

FIGURE 3. Open data categories by smart city dimension.

B. SMART ENVIRONMENT

Smart environment refers to all the tools and strategies put in place to ensure a sustainable environment. Cities account for 75% of global energy consumption and 80% of global CO2 emissions, which contributes to environmental degradation [57]. Thus, since a smart city must also be healthy

and sustainable [43], then environmental protection becomes imperative. In this light, Liu et al. [58] proposed a system for sharing geographic data in China. Thanks to the visualization of these data, it is possible to identify some environmental problems in Chinese cities such as the effects of population exposure to particulate matter (PM 2.5). Some researchers have also opted for ecology promotion [56] because of its important role in smart cities [59].

C. SMART PEOPLE

This field refers to several tools implemented to facilitate people's education. For example, many people do not know their cultures or are not interested in them. Thus, to promote art and culture, there are projects promoting their learning through digitalization such as ''Jecza Museum -the Sculpture Collection'' [60]. Based on open cultural and artistic data, this project aims to educate young people, with the purpose of conserving cultural heritage and art.

D. SMART GOVERNANCE

Smart governance is a democratic political governance framework that aims to improve government service delivery through digital technological innovations [61]. Indeed, sometimes people complain about government policy priorities, indicating that they are not too aligned with their own, or they also sometimes note a lack of transparency in the decision-making process. Thus, to address this problem, several works proposed to involve citizens more in the formulation of opinions and in political decision-making thanks to digital tools [61], [62]: this is electronic democracy. It improves governance by facilitating access to dialogue, and making decisions that are truly tailored to citizens in a transparent way; this will finally reinforce their trust in democratic procedures and values [63].

E. SMART ECONOMY

Smart economy refers to all technological solutions allowing businesses to be competitive in the market. In fact, there are several companies that offer the same services or products. This increases market concurrence. Faced with this competitiveness, it becomes costly to attract new customers [64] and even more costly to lose old ones. In fact, it is more profitable to succeed in keeping an existing clientele, since they are already satisfied with the services of a company and become the ambassadors of this company to others. Thus, identifying, in advance, the customers likely to unsubscribe, makes it possible to quickly put in place effective strategies to keep them loyal. In this light, Tang [64] proposed a model for predicting customer churn in the telecommunications industry, in order to allow mobile communication operators to maintain their competitiveness in the business market. One strategy for these telecom operators to maintain their clientele, for example, is to opt for intent-based systems [65] which are able to provide a good quality of experience.

The smart economy is also about finding a way to significantly reduce the costs due to consumption in a given city. In this context, Dizon and Pranggono [66] studied the city of Sheffield to reduce energy consumption due to street lights. Indeed, the lighting system in the city of Sheffield is conventional, i.e., the operation of the street lights is dependent on the solar clock. Using the StreetlightSim simulator and regional AADF (annual average daily traffic flow) data provided by Sheffield City Council, the authors studied this system and concluded that it is a very expensive energy system. In order to reduce the energy consumption of street lights, they studied four other approaches: the Chronosense scheme, Part Night scheme, Dynadimmer scheme, and Adaptive scheme. After the simulation, the Adaptive scheme showed better results, but unfortunately, these results are not reliable due to the lack of precision in some factors. Thus, the authors proposed to opt for the Dynadimmer scheme because it is the most reliable for the city of Sheffield and the most adapted to its streetlights, saving almost 50% of energy compared to the conventional scheme. The limitation of this study is therefore seen in the simulation of the adaptive scheme for the city of Sheffield. To provide reliable results, it needs a more accurate flow of traffic on the different roads of the city of Sheffield.

F. SMART LIVING

Smart living represents everything that allows people's quality of life. All of the above-mentioned domains contribute to smart living development. But beyond the services offered by these domains, smart living extends to other services such as health, entertainment, easy access, agriculture, etc.

1) SMART HEALTH

The healthcare field is not exempt from technological evolution. In fact, many hospitals are opting for the digitalization of their health systems, in order to increase the well-being of patients through more efficient diagnosis and monitoring, and also through remote access to their medical records called EHR (electronic health record). To this purpose, several works have approached this direction. For example, Chen et al. [67] proposed a mechanism for diabetes management able to predict with a high accuracy this disease, and then to give an adequate treatment to these diabetic patients. Liu et al. [68] studied people's physical activities to detect falls and psychiatric diseases. There is also cancer, which is a dangerous disease and one of the main causes of death [69]. This leads researchers to set up systems to detect it very early in order to treat it [70], [71], [72]. With the advent of COVID-19 causing a lot of health and economic damage, it has become more than necessary to forecast the number of infected cases, in order to help decision-makers to take the appropriate measures. In this context, several works have been carried out to accurately predict confirmed cases of covid [73], [74], [75], [76]. There is a more recent approach that is not focused on a specific disease but rather helps in

the general disease diagnosis [77]. In addition, it is essential to highlight that any change in these data causes false diagnoses and treatment that can lead to the patient's paralysis or death [78]. Moreover, some malicious people want to steal confidential patient information for resale. Given the two reasons mentioned, it is crucial to highly secure these EHR. Thus, Kumaar et al. [78] developed a hybrid intrusion detection system based on user profiling, to detect cyberattacks.

2) OTHERS

Regarding entertainment, there are, for example, streaming recommendation systems. These systems study the preferences and tastes of users in order to recommend movies that match them [79]. There is also the e-commerce sector that is developing. This one allows citizens to easily make their purchases online, saving time and energy, and avoiding certain diseases such as Covid-19. In this context, there is the largest online C2C (consumer-to-consumer) platform in China [80] which is based on user preferences to recommend suitable items.

G. REAL-TIME CASE STUDIES

1) A REAL-TIME DATA ANALYSIS FOR MEDICAL DIAGNOSIS USING FPGA-ACCELERATED NEURAL NETWORKS

Real-time analysis is really important because it allows more precision in the results. This is the case in the medical field where accuracy is very crucial since human life is at stake. Thus, it is necessary to rely not only on efficient data analysis algorithms but also to pay special attention to certain parameters such as the execution time of these algorithms. Sophisticated inference algorithms such as neural networks consume quite a lot of computing resources during their running, resulting in processing times that are contrary to the requirements of real-time. This problem is particularly observed with traditional processors such as the GPU (graphics processing unit) and CPU (central processing unit). Thus, in order to have an efficient real-time diagnostic system, Sanaullah et al. [81] proposed a real-time Multi-Layer Perceptron inference processor for medical diagnosis based on FPGA (field programmable gate arrays), that reduces latency thanks to correct parameter sizing. At the end of their experimentation on real-time cancer detection data, they concluded that their FGPA-based system outperforms the GPU and CPU. In their work, the authors showed how the variation of two parameters influences the performance of an application. Nevertheless, for an optimal solution, all contributions to latency must be taken into account.

2) A REAL-TIME EDGE COMPUTED ACCIDENT RISK INFERENCE SYSTEM

In order to improve road safety, Ovi et al. [82] proposed a real-time edge-computed accident risk inference system. Their system is based on a Deep Learning model combining MLP (multi-layer perceptron) and LSTM (long short-term memory) trained on various accident data collected from February 2016 to June 2020 in the US. To ensure their solution runs in real-time, they compressed their model and deployed it at the edge using an Intel NCS2 (neural compute stick 2) with the Raspberry Pi 4B. This technology not only allows them to achieve faster inference of their system but also allows low power and memory consumption. To show the performance of their model, they tested it in six US cities (Atlanta, Seattle, Detroit, Miami, Denver, and Chicago) and compared it to base models. Compared to the other models, the results showed that their solution is more accurate in predicting the risk of accidents. However, the accuracy of their prediction system in the city of Chicago is lower compared to that observed in the other five cities. Thus, further efforts are needed, to improve the accuracy of accident and non-accident risks in this city.

3) IoT-BASED PANEL FOR REAL-TIME TRAFFIC DATA MONITORING

Mora et al. [83] implemented a network of intelligent panels based on computer vision and IoT (internet of things), aiming to inform the driver about the traffic status in different locations around ZMG (Guadalajara Metropolitan Zone), allowing him/her to take the most convenient route. However, the authors implemented their system on the main avenue of ZMG which is a straight road. As a result, at the driver's level, their information panel displayed the traffic and travel time for two locations on this straight road. The fact that this information is not available on other roads around the driver is a limitation of this experiment because if for example there is a traffic jam, the driver does not have an optimal way to travel more fluidly. It is therefore necessary to add to this information panel a system on the driver's side, that automatically recommends a less time-consuming route in case of road disruption. In order to be effective, the system will have to rely on information from the surrounding panels and also from the driver's final destination, in order to determine the least time-consuming route.

Real-time implementation of research work is very important, as it allows us to really assess the feasibility of the proposed systems. However, this step is sometimes prevented by government regulations or insufficient resources.

IV. DATA ANALYTICS CONCEPT

The smart city services described above are implemented to meet the specific needs of the population or organizations. Identifying these needs is called profiling and it is possible thanks to careful data analytics. Thus, in this section, we will present different profiling types and models, and then we will detail the data analysis process. Finally, we will present a list of open data sources.

A. PROFILE MODELS AND TYPES

To really know a user and a city's daily life, it is necessary to know different profiles in terms of behavior, preferences, and intentions, thanks to a careful analysis of the collected

data. Thus, in order to carry out profiling, there are three main models to identify [29], [84]:

- behavior model: it is about identifying the different user habits, whether virtual or physical;
- interest model: the interest model is based on the user's preferences and tastes;
- intention model: the idea is to be able to predict what users want or will do, based on their behaviors and interests. In other words, knowing the purpose of their actions.

Based on these modeling processes, profiling can lead to the realization of a static profile or a dynamic profile. Eke et al. [29] defined a static profile as "a type of profile that maintains user information for a long period of time (. . .) like user's age and sex". Static profiling means analyzing static and predictable features of users [85]. Although some features are supposed to be fixed for a long time such as certain demographic information (e.g.: age), others can change at any time. For example, a user can declare a job and then, after 3 months, change to another job. In this case, static profiling is no longer appropriate, but dynamic profiling is used because it easily updates users' information based on the analysis of their actions in real time.

B. DATA PROCESSING CYCLE

1) DATA COLLECTION

We reiterate that user and city profiling requires data analytics. To this end, we list several types of data sources that can provide us with useful information on given users and a city.

a: CDR, GPS, SYSTEM LOGS, AND NETWORK TRAFFIC

Bianchi et al. [86] used CDR (call detail record) data provided by the telecom operator Orange for D4D (data for development) challenge. With these CDR data, the authors were able to collect certain details about the users' telephone exchanges such as the time and date of the exchanges, their duration, the latency time between two consecutive calls, the id of the prefecture to which the exchanges are attached, etc. These details allowed the authors, to highlight users' behavioral models. In addition to CDR data analysis, Vázquez-Salceda et al. [46] and Ayesha et al. [87] used users' GPS (global positioning system) data to obtain more accurate results on their spatial behaviors. Through the analysis of mobile users' call records, Garcia-Davalos and Garcia-Duque [88] discovered their social relationships i.e., their relatives. As for Lashkari et al. [30], in order to differentiate the normal user behavior from his/her abnormal behavior, they resorted to the system log analysis of this user. These logs make it possible to trace the user's activity history, thus serving to model his/her normal behavior. Still within the anomaly detection context, they also used the users' network traffic, like Akshay Kumaar et al. [78], whose analysis allows them to detect attacks such as DDoS. The network state is also an important data source in the effective implementation of an IDN (intent-driven network) [34] or IBN (intent-based network) [89], as it allows constant checks if the configurations have been well applied. The data cited above are usually provided by organizations.

b: SOCIAL NETWORK

Many people have at least one social account on the internet and there are enough who are active on their accounts. Thus, several research works opted for user profiling via the analysis of users' activities on social networks. While some researchers use it to improve recommendation systems in the tourism field [46], [48], others use it to identify certain users' features, such as their demographic information [90] or their tendency to volunteer [91]. To achieve their objectives, Farseev et al. [90] used data from Twitter, Instagram, Foursquare, and Facebook. Social networks are therefore very useful for user profiling. However, this source of data remains limited since many users still log in regularly, but stay discreet or do not reveal their real behavior in order to protect their privacy. They are not wrong to do that, as attackers often take advantage of shared information to clone users' identities for malicious purposes [30].

c: BROWSING HISTORY

The browsing history includes all the pages visited by users. It allows us to identify their browsing habits and therefore to deduce their interests. For this purpose, Lashkari et al. [30] used this data source for security purposes. Indeed, the aim was to analyze all the sites visited by the users and to identify potentially malicious sites. Thus, users who regularly have potentially malicious sites in their browsing history are considered to be users at risk.

These various online data are generally collected thanks to crawling techniques such as the use of APIs [88], [90], [91] and Bread-first-search method [91].

d: SMARTPHONE APPLICATIONS

Zhao et al. [31] gathered user profiling research works based on the analysis of smartphone app data. They classified these data into four categories: installed app lists, app installation behaviors, app metadata, and app usage records (app event logs). For example, Xie et al. [92] opted to analyze users' mobile healthcare application event logs, to detect scalpers of online healthcare services. It is also possible to obtain some contextual information about users such as their activities and location [51], [88]. These data from smartphone applications are generally collected through:

- APIs [88];
- specific applications such as AppSensor [93] and AppJoy [94] or even Futurefleet that is based on MCS (mobile crowdsourcing) [51];
- complete mobile sensing frameworks like Funf, Aware [95], Carat project [96], and Device Analyzer [97];
- organizations [86].

e: SURVEY DATA, GENERATED DATA, AND UNFIXED DATASET

Sometimes, researchers are faced with a lack of appropriate datasets for their work. Thus, to get better results, they can:

- collect data through wearable devices [68] or via a chatbot [98], both used by consenting people during a survey; in addition, there are sensors that allow for example to collect contextual data [46], [47] such as weather, traffic, etc.;
- generate data: Rajasinghe et al. [99] proposed a customizable software framework that can generate labeled network intrusion datasets on demand. This framework allows to have enough data to train an IDS (intrusion detection system) and consequently improve its performance;
- use an unfixed dataset to work. For example, Chen et al. [53] used a RL (reinforcement learning) algorithm called DQN (Deep Q Network), to solve a N-P hard problem of task offloading. Indeed, Reinforcement learning is a branch of machine learning that operates with dynamically collected data from the environment without using a static dataset.

2) DATA ANALYTICS WITH DATA MINING AND ML TECHNIQUES

Once data collection is done, it is time to move on to the data analytics step. This stage is very important because it allows us to deeply analyze the collected data, in order to discover correlations or patterns in these data, extract useful information and make predictions, thanks to data mining and ML techniques [100]. Sometimes these two concepts (data mining and ML) are confused, so what is the difference between them? In fact, they are complementary. Data mining requires human intervention to extract the most relevant information hidden in a dataset. To perform that task, it uses several disciplines such as statistics, database systems, pattern recognition, artificial intelligence, and machine learning algorithms (e.g., clustering, classification) [23]. Machine Learning is mainly about teaching a computer to learn and understand given parameters, in order to automate complex tasks without human intervention, like predictions [101]. Thus, ML can use valuable information from data mining as a resource to better learn the connections between relationships [101], [102].

The data analytics process follows this logic: service environment understanding, data understanding, data preparation, modeling, evaluation, and deployment [40].

a: SERVICE ENVIRONMENT AND DATA UNDERSTANDING

Service environment refers to the understanding reasons behind a service creation, to identify the key factors required to achieve the desired objectives. Data understanding consists of listing the various raw data collected, verifying their accuracy, and normalizing them when necessary. Normalizing data means transforming them into more understandable, uniform, and easy-to-process data, such as transforming geographic coordinates into addresses [88]. Kumaar et al. [78] normalized data using the standardized z-score formula to achieve faster and more accurate convergence of their machine learning model. Srinu et al. [103] used the technique of normalization of Min-Max scaling, to normalize input data into the interval (0-1).

b: DATA PREPARATION

➢ **Data Integration**

During the data preparation phase, several data sources can be merged to refine the analysis and make it more accurate. To this end, Atote et al. [104] proposed user profiling based on the synchronization of all the user's devices, in order to better identify his/her profile. Lashkari et al. [30] followed the same approach by combining four sources of user data to better monitor his/her activities on the internet. Sometimes, we notice users who give certain information on Facebook, but which are not on Twitter or LinkedIn for example. Thus, Song et al. [91] integrated all these data from several user's social networks, thanks to their MSNDC technique (multiple social network data completion). This technique consisted in finding the latent spaces shared by the different users' social networks, through the optimization of an objective function that consists in solving a decomposition of the NMF (non-negative matrix factorization) (with L1-norm regularization). In the same context, to integrate the different data, Farseev et al. [90] first tested two popular multimodal data fusion approaches: late fusion and early fusion. They then chose the late fusion technique because it gave them better results. Early fusion is a technique that concatenates multimodal features into a long feature vector, while late fusion integrates the results obtained by learning with each modality [105]. Having imbalanced data, Kumaar et al. [78] used the SMOTE algorithm to upsample the minority class data using the KNN (k-nearest neighbors) approach, in order to obtain an almost balanced dataset and so provide better model performance. The data obtained after this upsampling are in fact synthetic samples generated from the K nearest neighbors of each original data point. Abbasimehr et al. [75] used bootstrapping to generate new time series.

➢ **Data Cleaning**

The intention is to remove all data that are unnecessary for the objectives to be achieved. It is important because it permits to reduce the computation complexity [70]. In the context of location-based user profiling, Ayesha et al. [87] used a filter based on users' spatial behavior to identify load-sharing records. Their spatial activities were analyzed using Shannon entropy. Also, the Gabor filter was used during image preprocessing to reduce noise [72] and Pathan et al. [71] used the sharpening kernel technique to reduce image distortion. To extract users' tourism preferences from texts, Abbasi-Moud et al. [49] followed this text filtering process:

- part of speech tagging: identification of sentence elements (subject, verb, nouns, etc.);
- stop words elimination: eliminate words that have no specific meaning such as articles (e.g.: a, the);
- stemming: reduce the word to its word stem or the word root (e.g.: the stem of beating, beats, beaten is beat);
- extracting nouns: in a sentence, nouns are the most important words for efficient clustering [106].

To remove irrelevant data, Kumaar et al. [78] successively used the CFS (correlation-based feature selection) and recursive feature elimination techniques. Indeed, after combining two datasets with the same features, they used CFS to identify the strongly correlated features. And since two strongly correlated features will have the same effects on dependent variables, the idea was to remove one of them. Then, after calculating each remaining feature's significance using the p-value, they used a logistic regression classifier for a recursive feature elimination. Recently, Mohammed et al. [107] proposed a deep RL-based system to eliminate irrelevant IoT sensor data.

➢ **Dimension Reduction**

There are Machine Learning models whose performance degrades with too much data, such as KNN [40]. Another problem is that we cannot visualize the data with all dimensions, so we need to be able to display the most essential data. To solve these two problems, there are two popular techniques that reduce data without losing enough information: PCA (principal component analysis) [40], [86], [90] and LDA (linear discriminant analysis) [40], [108]. However, Amato and Lecce [109] proposed a new method called semi-pivoted QR approximation, to efficiently reduce a data set, and they demonstrated its performance compared to PCA.

➢ **Features Extraction Techniques**

Often, the collected data features are not all clearly defined. Thus, in order to extract really significant features, feature extraction techniques are used. Data from multiple sources are often multimodal. And each data mode has its own techniques for extracting features.

To identify users' habits from their CDR data, Bianchi et al. [86] established features based on observations and statistics. The features obtained are weekday, workday, day period, and previous calls. On another side, to discover the different individual mobility patterns, Ayesha et al. [87] selected the main features from application data, based on a correlation analysis technique. Garcia-Davalos and Garcia-Duque [88] used a statistical method to calculate the incoming and outgoing frequencies from different user call logs, in order to know the user's Personal Social Network. The selected features are call frequency for each hour, distance for each hour, appeared cell frequency, and appearance on weekdays or weekends. With a statistical method, some features can also be deduced heuristically, e.g., by counting the number of URLs, hashtags, slang words, emotion words, emoticons, etc. [90], [91]. To better understand visual user preferences, Farseev et al. [90] used LBP (local binary patterns), 64-D CH (color histogram), and Bag-of-visual-words as features. For the bag-of-visual-words, key points in each image were obtained by the difference of Gaussians technique, and their SIFT descriptors were extracted. Beyond that, to extract features on CT images, Vaiyapuri et al. [72] used MobileNet, and Wu et al. [110] proposed to use 3D-ResNet. Furthermore, in order to obtain a better prediction effect, a data decomposition method VMD (variational mode decomposition) has been proposed to reduce the complexity of the original data [111], [112]. This way can better determine the inherent data features. Thus, Li et al. [76] proposed to use GVMD (gradient-based optimizer variational mode decomposition) on covid-19 data, which is a better technique than the existing VMD.

c: MODELING AND EVALUATION

We will now move on to the data modeling and evaluation phases. In the modeling step, it is a question of representing the phenomenon behavior behind these data, in order to be able to solve other future problems.

➢ **NLP and Ontology Methods**

Computers can understand human language thanks to an artificial intelligence branch called NLP (natural language processing) [113]. Several techniques are used to perform NLP. LIWC (linguistic inquiry and word count) has been used to identify volunteer features [91] and also to determine the users' age and gender [90]. Indeed, LIWC is a tool that analyzes the linguistic words of a text, classifying them by psychological category of interest, in order to capture the users' emotions and personality. Also, it is possible to evaluate a word's importance in a text by calculating its frequency using the weight function TF-IDF (term frequency -inverse document frequency) [114]. There is also LDA (latent dirichlet allocation) which is a technique for modeling the document's subjects. More clearly, its objective is to find the subjects to which a document belongs, according to the words it contains. This helps to learn more about users' interests and personality [90], [91]. To develop an intent-based network, Jacobs et al. [98] used NER (named entity recognition) to extract and label entities from the operators' natural language intents and then used the Nile layer to translate them into network configuration commands. We can also semantically analyze a text using the FLM (fuzzy linguistic modeling) technique which identifies the meaning of the language used [40]. NLP has limitations in terms of accurate semantic understanding of a text. Thus, to accurately capture the meaning of a text, ontology is used [40]. In fact, an ontology is a ''conceptualization of a domain into a human-understandable, but the machine-readable format, which consists of entities, attributes, relationships, and axioms'' [115]. The ontology thus allows an unambiguous semantic analysis thanks to the different relations created between the different entities. So, in order to help doctors in their diabetes diagnostic decision-making, Chen et al. [67] developed an ontology-based model named OMDP. This model is indeed able to screen for diabetes and then provide appropriate treatment. To do so, they analyze patient data by combining several detailed pieces of knowledge about diabetes. Moreover, to understand text meaning, in order to infer users' tourism preferences, Abbasi-Moud et al. [49] opted for a semantic clustering score technique. This technique is based on a semantic similarity measure proposed by Wei et al. [116].

TABLE 3. Model evaluation metrics.

Evaluation metrics	Definitions
Precision [47], [48], [64], [68], [78], [98]	Rate of correct predictions among positive predictions.
Recall [47], [64], [68], [78], [98]	Allows us to know the percentage of positives well predicted by our model.
F-Measure [47], [64], [68], [78], [91], [98]	It is a synthesis of recall and accuracy metrics that is used to evaluate the ability of a model to correctly predict.
RMSE (root mean square error) and MAE (mean absolute error) [127], [128], [130]	The difference between the actual ratings and predicted ratings.
Hit-rate [48]	The success rate in accurate target shooting.
Accuracy [47], [64], [67], [77], [78], [91]	These are the correctly predicted data points compared to the data set.
ROC (receiver operator characteristic) curve [78]	A graph giving the true positive rate as a function of the false positive rate. A visual description of the ration of the performance results of Precision and Recall.
AUC (Area Under the Curve) [78]	AUC measures the probability that a random relevant item is ranked higher than a random irrelevant item.
P-value [87], [91]	Quantifies the risk of being wrong. It must be less than a statistical measure that represents the significance level of the tests.
Time Efficiency [77]	Time required to execute the model.
Sensitivity [67], [68]	The sensitivity measures the true positives rate.
Specificity [67], [68]	Specificity measures the true negative rate.
Coefficient of determination R^2 [128], [130]	It measures how well the collected data are explained by the prediction model.
SMAPE (symmetric mean absolute percentage error) [75]	This metric refers to an error percentage to evaluate the accuracy.

➢ **Statistics Methods**

TRProfiling: TRProfiling has been proposed [51], to provide itineraries that satisfy users' needs by taking into account several constraints. The users' travel needs are concretely described using statistical methods and are instantiated using ARIMA (autoregressive integrated moving average model). Then, to select the optimal route, it was first modeled as a TSP problem, and was solved using a heuristic algorithm named MCEA (multiconstraint evolution algorithm).

CityDNA: To better conceptualize a smart city, Moustaka et al. [117] opted for an approach that identifies the interrelations between smart city dimensions. Thus, inspired by the concepts of human DNA, they designed a CityDNA framework to represent urban profiles, in order to allow stakeholders to make suitable decisions for populations. To achieve this, CityDNA associated data collected on two city dimensions (smart mobility and smart economy) and used Pearson's Correlation Coefficient (PCC) to analyze the correlations between these dimensions' attributes. As a result, a double helix schema is generated and reflects the linear correlations or non-correlations between these attributes.

GCS-P Framework: In order to improve personalized location recommendations, Ma et al. [118] proposed a method that combines users' geographic, categorical and social preferences with location popularity using a linear fusion framework inspired by Ye et al. [119]. Geographic and social

TABLE 4. Open data sources.

Articles	Official data portals	Big data initiatives (obtained explicitly/implicitly)	Broader open data community (with sharing requirements)
[110] DeepMMSA: A Novel Multimodal Deep Learning Method for Non-small Cell Lung Cancer Survival Analysis	National Cancer Institute (NCI) has contracted with Washington University in Saint Louis to create The Cancer Imaging Archive (TCIA): https://cloud.google.com/healthcare-api/docs/resources/public-datasets/tcia?hl=fr	-	-
[75] A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting	Novel Coronavirus (COVID-19) Cases Data: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases	-	-
[80] Billion-scale commodity embedding for e-commerce recommendation in Alibaba	Amazon Electronics dataset: http://jmcauley.ucsd.edu/data/amazon/	Data collected from the mobile Taobao App	-
[73] Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization	Novel Coronavirus (COVID-19) Cases Data: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases	-	-
[123] CNN-Based Emotional Stress Classification using Smart Learning Dataset	WESAD dataset: https://github.com/WJMatthew/WESAD	-	-
[71] Breast Cancer Classification by Using Multi-Headed Convolutional Neural Network Modeling	Ultrasound dataset: https://scholar.cu.edu.eg/?q=afahmy/pages/dataset	-	-
[86] Identifying user habits through data mining on call data records	-	-	Orange Côte d'Ivoire published its users' phone dataset for its "Data for Development" (D4D) challenge.
[46] Making Smart Cities Smarter Using Artificial Intelligence Techniques for Smarter Mobility	-	Weather and traffic sensors data, GPS data. Also, social networks like Foursquare, Twitter, Instagram with crawling.	Data from mobile operator networks.
[87] User Localization Based on Call Detail Record	-	Data were collected via a mobile app for Android downloaded by users.	CDR data were provided by a regional ICT policy and regulatory think tank called LIRNEasia.
[88] User profile modelling based on mobile phone sensing and call logs	OSN (Open Storage Network) platforms.	Data were collected via mobile phone sensors.	-

TABLE 4. *(Continued.)* **Open data sources.**

[78] A Hybrid Framework for Intrusion Detection in Healthcare Systems Using Deep Learning	Canadian Institute for Cybersecurity platform: https://www.unb.ca/cic/datasets/ids-2017.html	-	-
[48] Efficient user profiling based intelligent travel recommender system for individual and group of users.	Yelp and TripAdvisor platforms.	-	-
[90] Harvesting multiple sources for user profile learning: A big data study	NUS-MSS: A Multi-Source Social Dataset from National University of Singapore: https://scholarbank.nus.edu.sg/handle/10635/137406	Foursquare, Facebook, Twitter with crawling.	-
[91] Multiple social network learning and its application in volunteerism tendency prediction	-	Twitter, Facebook, LinkedIn, About.me, Quora with crawling	-
[76] A new hybrid prediction model of cumulative COVID-19 confirmed data	Novel Coronavirus (COVID-19) Cases Data: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases	-	-
[92] User Profiling in Elderly Healthcare Services in China: Scalper Detection	-	-	Data provided by Qu Yi Yuan healthcare software company.
[51] A Shared Bus Profiling Scheme for Smart Cities Based on Heterogeneous Mobile Crowdsourced Data	-	Use of a Mobile Crowdsourced Data-based app.	-
[47] Travelers-Tracing and Mobility Profiling Using Machine Learning in Railway Systems	1. London Under-ground and Overground (LUO) dataset: https://data.london.gov.uk/dataset?tag=tube 2. UWB data: https://www.ofcom.org.uk/ 3. https://www.data.gov.uk/	-	-
[99] INSecS-DCS: A Highly Customizable Network Intrusion Dataset Creation Framework	-	-	Data set creation via software components available for public use under a MIT license.
[49] Tourism recommendation system based on semantic clustering and sentiment analysis	TripAdvisor platform	-	-
[67] OMDP: An ontology-based model for diagnosis and treatment of diabetes patients in remote healthcare systems	-	-	Data provided by hospital of Sun Yat-sen University
[64] Telecom customer churn prediction model combining k-means and xgboost algorithm	Customers churn data on Kaggle website	-	-

TABLE 4. *(Continued.)* **Open data sources.**

[61] E-democracy tools adoption: Experience of austria, croatia, italy, and slovenia	1. Open Data project in Italy: www.dati.gov.it 2. Slovenian Open Data Portal (OPSI): https://podatki.gov.si/data/	-	-
[125] Xgboost: a scalable tree boosting system	1. Allstate insurance claim dataset https://www.kaggle.com/c/ClaimPredictionChallenge 2. Higgs boson dataset: https://archive.ics.uci.edu/ml/datasets/HIGGS 3. Yahoo! learning to rank challenge dataset (Yahoo LTRC); Criteo terabyte click log dataset: https://labs.criteo.com/2013/12/download-terabyte-click-logs/	-	-
[77] Integrating medical code descriptions and building text classification models for diagnostic decision support.	Ambulatory Health Care Data: https://www.cdc.gov/nchs/ahcd/index.htm	-	-
[127] Digital Twin Mobility Profiling: A Spatio-Temporal Graph Learning Approach; Digital Twin Mobility Profiling: A Spatio-Temporal Graph Learning Approach	-	-	The Huaian dataset generated by Panda Bus Company, China
[128] Pedestrian flow prediction in open public places using graph convolutional network	-	Pedestrian flow data collected from detectors. Also, the weather and temperature data were collected from the Meteorological Bureau of Shenzhen Municipal website.	-
[129] Enhancing pedestrian mobility in smart cities using big data	Data were obtained via the City of Melbourne Open Data Platform: https://data.melbourne.vic.gov.au/	-	-
[79] A hybrid recommender system for recommending relevant movies using an expert system	MovieLens dataset: https://www.kaggle.com/datasets/vedapragnareddy/movie-lens	-	-
[130] Forecasting of bicycle and pedestrian traffic using flexible and efficient hybrid deep learning: a survey approach.	-	-	Datasets created and maintained by the Seattle Department of Transportation in the USA.
[118] Location recommendation by combining geographical, categorical, and social preferences with location popularity	1. Weeplaces dataset: http://www.yongliu.org/datasets/ 2.Yelp dataset: https://www.yelp.com/dataset	-	-

TABLE 4. *(Continued.)* **Open data sources.**

[50] Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach	Helsinki's journey planner: https://hri.fi/en_gb/	-	-
[58] Understanding urban china with open data	1. Beijing City Lab: https://www.beijingcitylab.com/data-released-1/ 2. New York City (NYC) platform: https://nycopendata.socrata.com/ 3. Peking University – Lincoln Institute Center: https://www.lincolninst.edu/research-data 4. VGI mapping platforms like OpenStreetMap 5. Location-enabled social media applications like Flickr	-	-
[56] Eco-efficient mobility in smart city scenarios	-	The data were retrieved from the OBU's embedded sensors	-
[74] The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method	Data collected by John Hopkins University: https://github.com/CSSEGISandData/COVID-19	-	-
[109] Data Analysis for Information Discovery	http://archive.ics.uci.edu/ml	-	-
[81] Real-time data analysis for medical diagnosis using FPGA-accelerated neural networks	home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp	-	-
[82] ARIS: A Real Time Edge Computed Accident Risk Inference System	https://smoosavi. org/datasets/us_accidents Created by S. Moosavi, M. H. Samavatian, S. Parthasarathy, and R. Ramnath	-	-
[66] Smart streetlights in Smart City: a case study of Sheffield	https://www.gov.uk/government/statistical-data-sets/road-traffic-statistics-tra	-	-

preferences were computed using the power law function and categorical preferences are computed based on the semantic similarity between location tags.

MSNL (Multiple Social Network Learning): Besides recommendation services, websites can help to discover certain demographic information. To this end, Song et al. [91] implemented an analytical solution allowing to identify volunteers on social networks. This solution consisted of studying a linear mapping function for each social network, and then establishing the final model as a linear combination of these different functions trained on the social networks concerned. They first used the least square loss on the mapping function to obtain an objective function. And unlike some previous analytical proposals (iSFS [120] and regMVMT [121]), the authors have regularized the objective function by considering the trust and coherence parameters of social networks. The resolution of these two parameters was done using an optimization of the objective function, by fixing one of them at each iteration.

➢ **Machine Learning**

Classification: After selecting the features as mentioned in the data preparation section, Ayesha et al. [87] proceeded to

user profiling in order to identify their localization. Due to the class imbalance problem in the mobile application and CDR datasets, SVMwc (support vector machine weighted classes) model was chosen to classify data. To obtain the users' age and gender, Farseev et al. [90] instead applied RF (random forest) model to each data type of three social networks. They finally integrated the different RF classifiers into their global learning model based on a SHCR (stochastic hill climbing with random restart) optimization. Menon et al. [108] also proposed an approach based on the RF classifier, to regularly inspect the health status of diabetics. However, they improved the basic RF by integrating feature selection weighting during classification. Their approach is called ASV-RF (advanced-spatial-vector-based random forest). From a marketing side, it is very important for telecom operators to predict the users' churn rate in order to prevent it. Thus, Tang [64] proposed a churn prediction model. Indeed, with the aim of improving the forecast accuracy and generalization ability, after performing clustering on the training data, the author applied the XGBoost algorithm to classify the obtained clusters.

Clustering: In order to learn users' habits from their CDR data, Bianchi et al. [86] proposed a multi-agent LD-ABCD algorithm that relies on a suitable configuration parameters list to return meta-clusters describing these users' habits. This parameter list defined the features by which the elements of the meta clusters are similar to each other thanks to a dissimilarity measure. They showed that their clustering approach is stable and complete. To achieve its goals, LD-ABCD used multiple Markovian walks to discover small-radius meta-clusters, set a threshold to retain only quality ones, and used Boolean vectors to distribute the CDR data among these meta-clusters. Unlike Bianchi et al. [86], Ayesha et al. [87] did not just stop at modeling the user CDR data. They also applied a filter based on a user's spatial behavior (Shannon entropy), using GPS data, to solve a main limitation of CDR data which is load-sharing records. Clearly, a user A can call being in the coverage area of a cell tower X, but since X is not available, his/her call is attached to an adjacent available tower Y. The problem is that the Y location does not reflect the real position of user A. Based on this filter of users' spatial behaviors, the authors used DBSCAN (density-based spatial clustering of applications with noise) clustering to group the CDR data, and to identify the different stay regions of these users. It is true that, after the spatial filter is applied, the different cell towers to which user A has been attached during his/her calls are known, but this is still insufficient. Indeed, the exact location of user A is needed and this location corresponds to the centroid of the different cell towers that have been assigned to him/her. Hence, in addition to the load sharing record parameter, the authors added two other parameters to the clustering: signal strength of the particular cell tower, and the number of days that appeared in the particular cell tower. The centroid was obtained using weighted k-means++ algorithm.

Moreover, Garcia-Davalos et al. [88] used a Jenks Natural Breaks partitioning clustering algorithm to obtain cocentric circles of user's personal social network, based on incoming and outgoing frequencies that were computed from users' call logs. The objective was to identify the users' different social relationships such as their affinity group, sympathy group, etc. Then, based on this personal social network and users' location data, they inferred the most common situations in which users find themselves, using the statistical technique of Bayesian inference. Furthermore, to infer individual and collective user activity-behavior profiles, Logesh et al. [48] applied the Fuzzy C-means clustering algorithm on user demographic and preference data. User preferences were obtained by the proposed UUPM technique which combined contextual preference mining, sentiment analysis and UTA (utility theory additive) method. Finally, several POIs (point of interest) filters were applied on the obtained users' profiles, in order to recommend them a personalized list of top N POIs.

TABLE 5. Open data for traffic prediction problems [39].

Data type	Dataset name
Traffic Sensor Data	METR-LA, Performance Measurement System (PeMS) Data
Taxi Data	T-drive, SHSpeed (Shanghai Traffic Speed), TaxiBJ, TaxiSZ, TaxiCD, TaxiNYC
Ride-hailing Data	UberNYC, Didi GAIA Open Data
Bike Data	BikeNYC, BikeDC, BikeChicago
Subway Data	SHMetro, HZMetro

Neural Networks: In 2020, a review gathered works using deep learning to realize smart city applications focused on mobility, education, urbanization, security, and health [18]. Other interesting and recent works have been published in these different application fields. To protect patients' medical data (EHR) from any cyberattack, Kumaar et al. [78] implemented a deep learning network called ImmuneNet that analyzes network traffic for any intrusion into the EHR system. Immunet is a sequence of blocks whose result is output from a residual operation between the first and the last block. And each block contains simple linear projections followed by mish activation and layer normalization. Still in the medical field, a method to know tumor's limits or a tissue's volume is image segmentation which consists in analyzing an image by associating to each pixel a label (semantic segmentation), and by delimiting each interest object on the image (instance segmentation) [122]. An efficient technique to perform this segmentation is deep learning [122]. Vaiyapuri et al. [72] proposed an EPO-MLT (emperor penguin optimizer with multilevel thresholding), for pancreatic tumor segmentation on CT images. Then, to classify tumors, an AE (auto encoder) was used with a MLO (multileader optimization) technique to fine-tune these AE model parameters. Also, Pathan et al. [71] proposed a model based on a multi-headed CNN (convolution neural network) to recognize breast cancer. To effectively classify emotional stress features from vital signs, Andreas et al. [123] used a CNN in conjunction with OTL (online transfer learning).

In fact, transfer learning is a ML type that is used to transfer knowledge from a model trained on a source dataset to a model to be trained on a more specific and reduced target dataset [124]. To help clinicians in making diagnostic decisions, a system based on medical code descriptions has been proposed [77]. Indeed, several patient symptoms are recorded as medical codes and diagnostics are performed directly on these codes using, for example, XGBoost [125] and MLP [126]. However, Tang et al. [77] demonstrated that using the textual description of these medical codes increases diagnostic accuracy. Thus, after mapping each medical code to its textual description, the authors performed NLP using three separate convolutional neural network models for text, TextCNN, TextRCNN, and CharCNN, to evaluate the accuracy and time efficiency. At the end of this experiment, each of these three models on textual descriptions presented better results than using XGBoost and MLP directly on medical codes.

To obtain an accurate user mobility profile, Chen et al. [127] recently proposed a system called Digital Twin Mobility Profiling. First, the Digital Twin technology was used to create a virtual copy of a physical traffic network, in order to simulate several scenarios without disturbing the physical network [127]. Then, to learn node profiles on this DT network, DACN (dilated alignment convolution network) and TCN (temporal convolution network) were used in parallel to represent and analyze fine-grained spatio-temporal interaction in the network. In the same period, Liu et al. [128] focused on pedestrian traffic specifically, in order to guarantee the road safety of pedestrians in public spaces through adapted urban configurations. Indeed, unlike Carter et al. [129] that were limited to the exploration and visualization of pedestrian traffic using Microsoft Excel and PowerBI, Liu et al. [128] extended to crowd flow prediction in the streets. For this purpose, they proposed a model based on the GCN (graph convolutional network). Likewise, there is a recent survey that gathered traffic forecasting works also based on graph neural network [39]. Harrou et al. [130] proposed an approach that predicts not only pedestrian traffic but also cyclist traffic. This approach is based on a guided-attention hybrid deep learning architecture called GAHD-VAE. Indeed, to obtain better predictions, two main techniques have been integrated into the VAE (variational autoencoder): a self-attention mechanism was used to know the most relevant part of features, while the recurrent neural network LSTM (long short-term memory) served to efficiently model the temporal dependencies in time series data.

We add to this set of presented works two other surveys which also regroup other works based on the ML [21], [22].

➢ **Hybrids Methods**

Although statistical and machine learning models are efficient, some researchers opt for hybrid models combining statistics and machine learning to get more accurate models. In this context, Abbasimehr and Paki [73] proposed three hybrid models based on BO (bayesian optimization), to predict COVID-19 infected cases. BO has been used to efficiently determine the optimal hyperparameters of different DL (deep learning) models. In fact, its efficiency and superiority over grid search has been demonstrated [111], [131]. Thus, the first one combined multi-head attention and BO, the second one combined LSTM and BO, and the third one combined CNN and BO. After comparison, the results concluded that the model based on LSTM and BO was better for long-horizon forecasting. Besides, a hybrid model based on a DL LSTM and a statistical Markov method has been proposed [74]. LSTM was used to predict the cumulative number of infected cases, while Markov was used to correct the prediction error of the LSTM model. After efficiently decomposing the data using their GVMD method, Li et al. [76] processed them with a model combining ELM (extreme learning machine) and ARIMA.

On the IoV side, the most suitable machine learning for task processing is reinforcement learning. Indeed, the advent of IoV has led to the creation of several intelligent applications such as autonomous driving, video-aided real-time navigation, etc. [53]. Since vehicular resources are insufficient to guarantee a good quality of experience for these applications, task processing is performed by MEC (multi-access edge computing) servers. However, MEC servers also suffer from insufficient computation services. In addressing this problem, Chen et al. [53] proposed a distributed computation offloading strategy in IoV. The first step was to model the optimal offloading problem as a Markov decision process. Then, they used the Deep Q Network to determine the optimal offloading and task allocation scheme.

After modeling, the next step is evaluation. This step allows us to evaluate how well the model created meets the identified needs. The evaluation is done by testing the model on real data to identify errors. We have grouped in Table 3 the metrics commonly used in scientific articles.

Fig. 4 summarizes the data mining and ML techniques presented in this article. These techniques were used since 2015.

C. DATA AVAILABILITY

Open data can be accessed through three main ways which are: official data portals (via the internet), big data initiatives (obtained explicitly or implicitly via crawling techniques), and the broader open data community (with sharing requirements. e.g.: sharing for research purposes only) [5], [58]. To facilitate the search for useful data for future scientific works, we gather in Table 4 some data sources specified by the presented works.

In addition to the different data sources that we have summarized in Table 4, we introduce other open datasets from a survey about traffic forecasting [39] (Table 5).

V. RESEARCH DIRECTIONS

A. SOME UNDER-EXPLOITED ASPECTS OF SMART CITY

1) SMART HEALTH

Most of the solutions proposed at the moment target a specific disease. It would be interesting to develop a more global

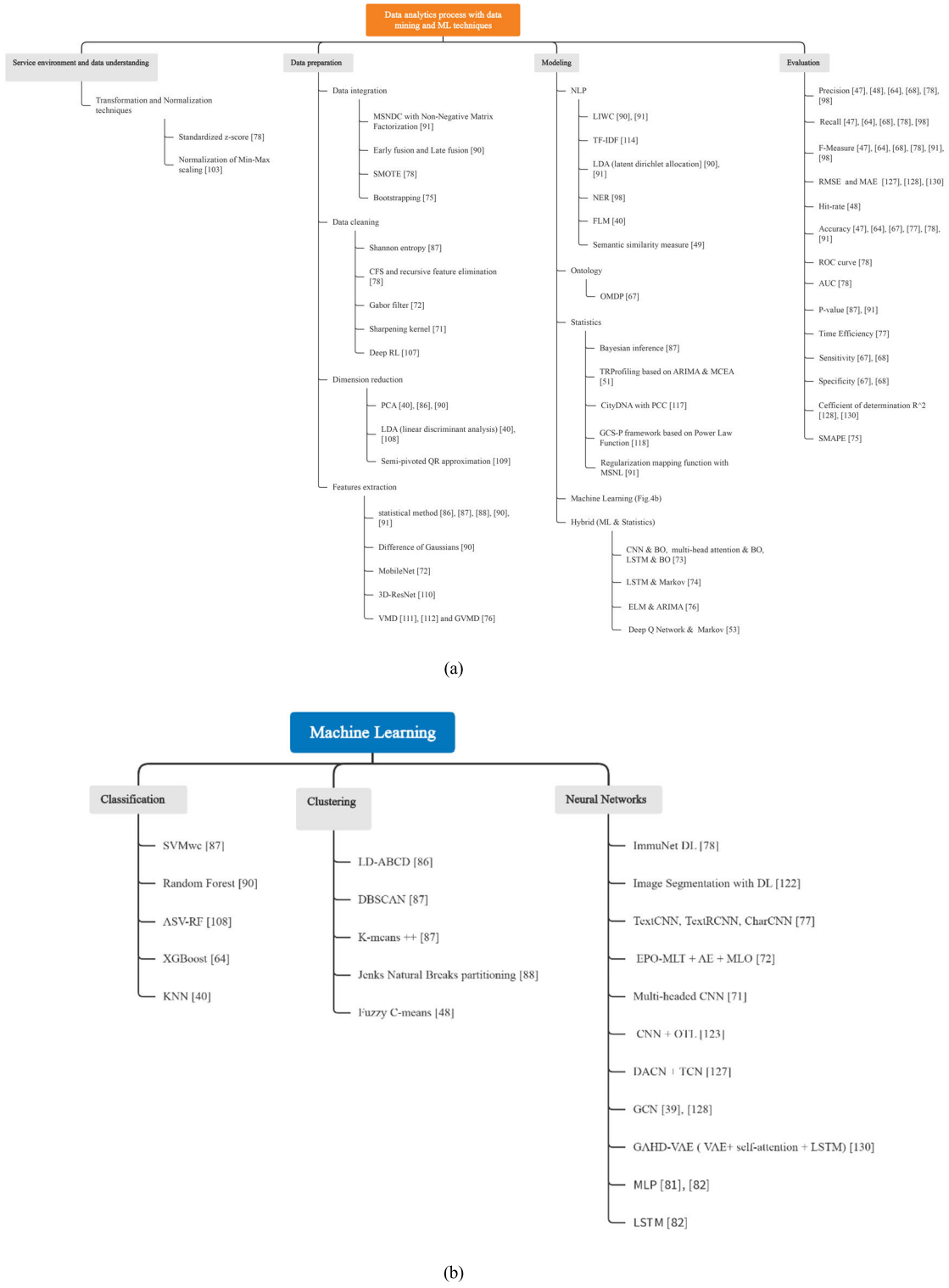

FIGURE 4. Data analytics: (a) data analytics process with data mining and ML techniques, and (b) ML techniques.

system allowing to predict exactly the rate of occurrence of each disease for a given patient as well as the date when the disease might occur.

2) SMART ECONOMY

Intent-based networks are very useful for companies, especially for telecom operators that are in a perpetual need

to improve the QoE (quality of experience) of their users in order to retain them. These intent-based networks, with natural language, can indeed quickly handle a wide range of network policies.

3) SMART MOBILITY

It is necessary to deploy in real transport systems the smart mobility solutions that take into account spatio-temporal relationships.

4) INTEROPERABILITY OF SMART CITY DIMENSIONS

It is necessary to develop smart city applications taking into account interrelations between its six dimensions (smart mobility, smart environment, smart economy, smart people, smart governance, smart living).

B. HOW TO OVERCOME THE LACK OF DATA?

Smart city applications still face a lack of data. Several directions are to be explored to address this problem.

1) PARSIMONIOUS NEURAL NETWORKS

With a parsimony-based approach, it is possible to obtain valuable information from little data, while significantly reducing computational and time costs. Thus, it is an area to be further explored to overcome the problem of limited data.

2) TRANSFER LEARNING

Another technique is to use Transfer Learning, which transfers knowledge from a model trained on a source dataset to a model that is to be trained on more specific and reduced target data. This approach is appropriate when there are already models trained on large datasets for a more general application, while our application is about the specificity of this general context.

The main cause of this data scarcity is mainly the data security aspect. Therefore, it is necessary to make sure to process these data, while respecting the rules regarding the protection of users' privacy established by the GDPR. In this context, there are:

3) FEDERATED LEARNING (FL)

This approach mainly addresses the need to obtain accurate ML models while avoiding the sharing of raw user data. To do so, the ML model is trained on local data at each stakeholder level, and then each resulting gradient list is sent to a central server for aggregation, to build a more robust model. In the end, the parameters of this new global model are shared among the different stakeholders. To secure access to these parameters, a method is to use the blockchain in conjunction. This FL and Blockchain-based approach can be further explored for user profiling applications requiring very sensitive information about users such as their political, religious, sexual affiliation, etc.

4) PROMOTE DATA SHARING BY ENSURING THE SECURITY AND ETHICAL USE OF DATA

The FL and Blockchain based approach shows promise in ensuring user data privacy by avoiding real data sharing. However, there are applications that require on the one hand exploring the raw data to detect any potential hidden information; and on the other hand, that require data traceability in order to understand the background of a given phenomenon to avoid it in the future. This is the case of natural disaster management applications through the prediction of potential pandemics such as covid-19 or the prediction of earthquakes such as the one of magnitude 7.8 that occurred in Turkey and Syria. This type of disaster management allows to avoid many human casualties.

Therefore, another area to explore is using the blockchain to ensure the secure sharing of reliable data between different stakeholders. Thus, with blockchain, we will no longer have to worry about the reliability of the data sources that impact the results of the prediction model. Once the sharing is complete, it will be possible to use these heterogeneous and reliable data to develop even more accurate and robust ML models. Besides, in this blockchain, it would be reasonable for each user to have access to his/her data and to be aware of the processing performed, in order to be reassured about the ethical data use.

VI. CONCLUSION

The survey's purpose was to show the link between open data and the smart city in all aspects. To do this, we first conducted a comparative study of some existing surveys between 2015 and 2023, in order to identify the gaps to be addressed. As a limitation, we note that the existing surveys do not cover all aspects of the open data concept in the smart city. Therefore, our survey combines ''smart city applications'', ''open data categories adapted to the smart city'', ''data analytics process'', ''data mining and ML techniques'', ''open data sources'' and gives an overview of existing surveys. We clarified the difference between big data and open data, and then between data mining and ML. We have further detailed the data analytics process from collection to evaluation, presenting existing techniques since 2015 by analysis step and data type. Each solution presented, based on open data analytics, contributes effectively to smart city development and to quality-of-life improvement for populations. However, even if open data have enormous advantages for smart city development, some challenges remain. Thus, to overcome them, we have presented some future research directions.

REFERENCES

[1] M. Batty, *The New Science of Cities*. London, U.K.: MIT Press, 2013.
[2] D. Ahlers. (2020). *Making Sense of the Urban Future: Recommendation Systems in Smart Cities*. [Online]. Available: https://ceur-ws.org/Vol-2697/paper5_complexrec.pdf
[3] Smart City, ''How does data sharing contribute to better living in our daily urban life? 'comment le partage de données contribute-t-il à mieux vivre notre quotidien urbain?'' in *Proc. BNP Paribas Workshop Forum of Avignon*, 2014, pp. 1–5. [Online]. Available: https://www.forum-avignon.org/sites/default/files/editeur/Atelier_BNP_Paribas_position_paper_Session_4_Smart_Cities_VF_1709.pdf
[4] Y. Ma, G. Li, H. Xie, and H. Zhang, ''City profile: Using smart data to create digital urban spaces,'' in *Proc. 3rd Int. Conf. Smart Data Smart Cities*, Delft, Netherlands, Oct. 2018, pp. 75–82.

[5] J. Gurin, *Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation*, New York, NY, USA: McGraw-Hill, 2014.

[6] A. E. Prieto, J. Mazón, and A. Lozano-Tello, "Framework for prioritization of open data publication: An application to smart cities," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 1, pp. 131–143, Jan. 2021, doi: 10.1109/TETC.2019.2893016.

[7] L. Anthopoulos, "Understanding the smart city domain: A literature review," in *Public Administration and Information Technology*, vol. 8. Cham, Switzerland: Springer, 2015, pp. 9–21.

[8] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustain. Cities Soc.*, vol. 38, pp. 697–713, Apr. 2018, doi: 10.1016/j.scs.2018.01.053.

[9] A. Camero and E. Alba, "Smart city and information technology: A review," *Cities*, vol. 93, pp. 84–94, Oct. 2019, doi: 10.1016/j.cities.2019.04.014.

[10] J. Stübinger and L. Schneider, "Understanding smart city—A data-driven literature review," *Sustainability*, vol. 12, no. 20, pp. 1–23, 2020, doi: 10.3390/su12208460.

[11] A. M. Shahat Osman and A. Elragal, "Smart cities and big data analytics: A data-driven decision-making use case," *Smart Cities*, vol. 4, no. 1, pp. 286–313, Feb. 2021, doi: 10.3390/smartcities4010018.

[12] V. Grossi, F. Giannotti, D. Pedreschi, P. Manghi, P. Pagano, and M. Assante, "Data science: A game changer for science and innovation," *Int. J. Data Sci. Analytics*, vol. 11, no. 4, pp. 263–278, May 2021, doi: 10.1007/s41060-020-00240-2.

[13] J. Attard, F. Orlandi, S. Scerri, and S. Auer, "A systematic review of open government data initiatives," *Government Inf. Quart.*, vol. 32, no. 4, pp. 399–418, Oct. 2015, doi: 10.1016/j.giq.2015.07.006.

[14] Y. Gao, M. Janssen, and C. Zhang, "Understanding the evolution of open government data research: Towards open data sustainability and smartness," *Int. Rev. Administ. Sci.*, vol. 89, no. 1, pp. 59–75, Mar. 2023, doi: 10.1177/00208523211009955.

[15] M. Buchinger, P. Kuhn, A. Kalogeropoulos, and D. Balta, "Towards interoperability of smart city data platforms," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2021, pp. 1–6.

[16] F. T. Neves, M. de Castro Neto, and M. Aparicio, "The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring," *Cities*, vol. 106, Nov. 2020, Art. no. 102860, doi: 10.1016/j.cities.2020.102860.

[17] J. Souza, A. Francisco, C. Piekarski, and G. Prado, "Data mining and machine learning to promote smart cities: A systematic review from 2000 to 2018," *Sustainability*, vol. 11, no. 4, p. 1077, Feb. 2019, doi: 10.3390/su11041077.

[18] S. Bhattacharya, S. R. K. Somayaji, T. R. Gadekallu, M. Alazab, and P. K. R. Maddikunta, "A review on deep learning for future smart cities," *Internet Technol. Lett.*, vol. 5, no. 1, p. e187, Jan. 2022, doi: 10.1002/itl2.187.

[19] D. Iskandaryan, F. Ramos, and S. Trilles, "Air quality prediction in smart cities using machine learning technologies based on sensor data: A review," *Appl. Sci.*, vol. 10, no. 7, p. 2401, Apr. 2020, doi: 10.3390/app10072401.

[20] V. Moustaka, A. Vakali, and L. G. Anthopoulos, "A systematic review for smart city data analytics," *ACM Comput. Surveys*, vol. 51, no. 5, pp. 1–41, Sep. 2019, doi: 10.1145/3239566.

[21] L. Hurbean, D. Danaiata, F. Militaru, A.-M. Dodea, and A.-M. Negovan, "Open data based machine learning applications in smart cities: A systematic literature review," *Electronics*, vol. 10, no. 23, p. 2997, Dec. 2021, doi: 10.3390/electronics10232997.

[22] I. H. Sarker, "Smart city data science: Towards data-driven smart cities with open research issues," *Internet Things*, vol. 19, Aug. 2022, Art. no. 100528, doi: 10.1016/j.iot.2022.100528.

[23] A. Lausch, A. Schmidt, and L. Tischendorf, "Data mining and linked open data–new perspectives for data analysis in environmental research," *Ecological Model.*, vol. 295, pp. 5–17, Jan. 2015, doi: 10.1016/j.ecolmodel.2014.09.018.

[24] J. Peng, K.-K.-R. Choo, and H. Ashman, "User profiling in intrusion detection: A review," *J. Netw. Comput. Appl.*, vol. 72, pp. 14–27, Sep. 2016, doi: 10.1016/j.jnca.2016.06.012.

[25] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and privacy in smart city applications: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 122–129, Jan. 2017, doi: 10.1109/MCOM.2017.1600267CM.

[26] J. López-Quiles and M. R. Bolívar, "Smart technologies for smart governments: A review of technological tools in smart cities," *Public Admin. Inf. Technol.*, vol. 24, pp. 1–18, Jul. 2018, doi: 10.1007/978-3-319-58577-2_1.

[27] A. H. Alavi, P. Jiao, W. G. Buttlar, and N. Lajnef, "Internet of Things-enabled smart cities: State-of-the-art and future trends," *Measurement*, vol. 129, pp. 589–606, Dec. 2018, doi: 10.1016/j.measurement.2018.07.067.

[28] R. W. S. Ruhlandt, "The governance of smart cities: A systematic literature review," *Cities*, vol. 81, pp. 1–23, Nov. 2018, doi: 10.1016/j.cities.2018.02.014.

[29] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "A survey of user profiling: State-of-the-art, challenges, and solutions," *IEEE Access*, vol. 7, pp. 144907–144924, 2019, doi: 10.1109/ACCESS.2019.2944243.

[30] R. H. Lashkari, M. Chen, and A. A. Ghorbani, "A survey on user profiling model for anomaly detection in cyberspace," *J. Cyber Secur. Mobility*, vol. 8, no. 1, pp. 75–112, 2018, doi: 10.13052/jcsm2245-1439.814.

[31] S. Zhao, S. Li, J. Ramos, Z. Luo, Z. Jiang, A. K. Dey, and G. Pan, "User profiling from their use of smartphone applications: A survey," *Pervas. Mobile Comput.*, vol. 59, Oct. 2019, Art. no. 101052, doi: 10.1016/j.pmcj.2019.101052.

[32] T. Aljowder, M. Ali, and S. Kurnia, "Systematic literature review of the smart city maturity model," in *Proc. Int. Conf. Innov. Intell. Informat., Comput., Technol. (ICT)*, Sakhier, Bahrain, Sep. 2019, pp. 22–23.

[33] K. Soomro, M. N. M. Bhutta, Z. Khan, and M. A. Tahir, "Smart city big data analytics: An advanced review," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 9, no. 5, p. e1319, 2019, doi: 10.1002/widm.1319.

[34] L. Pang, C. Yang, D. Chen, Y. Song, and M. Guizani, "A survey on intent-driven networks," *IEEE Access*, vol. 8, pp. 22862–22873, 2020, doi: 10.1109/ACCESS.2020.2969208.

[35] A. Heras, A. Luque-Sendra, and F. Zamora-Polo, "Machine learning technologies for sustainability in smart cities in the post-COVID era," *Sustainability*, vol. 12, no. 22, pp. 1–25, 2020, doi: 10.3390/su12229320.

[36] C. Avci, B. Tekinerdogan, and I. N. Athanasiadis, "Software architectures for big data: A systematic literature review," *Big Data Analytics*, vol. 5, no. 1, p. 5, Dec. 2020, doi: 10.1186/s41044-020-00045-1.

[37] T. Yigitcanlar, K. Desouza, L. Butler, and F. Roozkhosh, "Contributions and risks of artificial intelligence (AI) in building smarter cities: Insights from a systematic review of the literature," *Energies*, vol. 13, no. 6, p. 1473, Mar. 2020, doi: 10.3390/en13061473.

[38] T. M. Ghazal, M. K. Hasan, M. T. Alshurideh, H. M. Alzoubi, M. Ahmad, S. S. Akbar, B. Al Kurdi, and I. A. Akour, "IoT for smart cities: Machine learning approaches in smart healthcare—A review," *Future Internet*, vol. 13, no. 8, p. 218, Aug. 2021, doi: 10.3390/fi13080218.

[39] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," *Exp. Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 117921, doi: 10.1016/j.eswa.2022.117921.

[40] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: Recommendation models, techniques, and application fields," *Electronics*, vol. 11, no. 1, p. 141, Jan. 2022.

[41] I. Vardopoulos, M. Papoui-Evangelou, B. Nosova, and L. Salvati, "Smart 'tourist cities' revisited: Culture-led urban sustainability and the global real estate market," *Sustainability*, vol. 15, no. 5, p. 4313, Feb. 2023, doi: 10.3390/su15054313.

[42] S. Cammers-Goodwin, "Open data insights from a smart bridge datathon: A multi-stakeholder observation of smart city open data in practice," *Smart Cities*, vol. 6, no. 2, pp. 676–691, Feb. 2023, doi: 10.3390/smartcities6020032.

[43] (Jun. 2016). *Recommendation ITU-T Y.4900/L.1600*. [Online]. Available: : https://www.itu.int/en/ITU-T/ssc/Pages/info-ssc.aspx

[44] R. Giffinger and H. Gudrun, "Smart cities ranking: An effective instrument for the positioning of the cities?" *ACE: Archit., City Environ.*, vol. 4, no. 12, pp. 7–26, Feb. 2010.

[45] J. Gurin. *Big Data and Open Data: What's What and Why Does it Matter?* Accessed: Feb. 2023. [Online]. Available: https://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government

[46] J. Vázquez-Salceda, S. Alvarez-Napagao, A. Tejeda-Gómez, L. Oliva, D. Garcia-Gasulla, I. Gómez-Sebastía, and V. Codina, "Making smart cities smarter using artificial intelligence techniques for smarter mobility," in *Proc. Int. Conf. Smartgreens*. Barcelona, Spain: SciTePress, 2014, pp. 7–11.

[47] S. M. Asad, K. Dashtipour, S. Hussain, Q. H. Abbasi, and M. A. Imran, "Travelers-tracing and mobility profiling using machine learning in railway systems," in *Proc. Int. Conf. U. K.-China Emerg. Technol. (UCET)*, Glasgow, U.K., Aug. 2020, pp. 1–4.

[48] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, and X. Li, "Efficient user profiling based intelligent travel recommender system for individual and group of users," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 1018–1033, Jun. 2019, doi: 10.1007/s11036-018-1059-2.

[49] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Exp. Syst. Appl.*, vol. 167, Apr. 2021, Art. no. 114324, doi: 10.1016/j.eswa.2020.114324.

[50] S. Jäppinen, T. Toivonen, and M. Salonen, "Modelling the potential effect of shared bicycles on public transport travel times in greater helsinki: An open data approach," *Appl. Geography*, vol. 43, pp. 13–24, Sep. 2013, doi: 10.1016/j.apgeog.2013.05.010.

[51] X. Kong, F. Xia, J. Li, M. Hou, M. Li, and Y. Xiang, "A shared bus profiling scheme for smart cities based on heterogeneous mobile crowdsourced data," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1436–1444, Feb. 2020, doi: 10.1109/TII.2019.2947063.

[52] F. Rasheed Lone, H. Kumar Verma, and K. Pal Sharma, "Evolution of VANETS to IoV," *Tehnički glasnik*, vol. 15, no. 1, pp. 143–149, Mar. 2021, doi: 10.31803/tg-20210205104516.

[53] C. Chen, Z. Wang, Q. Pei, C. He, and Z. Dou, "Distributed computation offloading using deep reinforcement learning in Internet of Vehicles," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2020, pp. 823–828.

[54] J. Teter, "Transport," Int. Energy Agency, Paris, France, Sep. 2022. Accessed: Jun. 2023. [Online]. Available: https://www.iea.org/reports/transport

[55] R. Vidhi, P. Shrivastava, and A. Parikh, "Social and technological impact of businesses surrounding electric vehicles," *Clean Technol.*, vol. 3, no. 1, pp. 81–97, Feb. 2021.

[56] R. Sanchez-Iborra, L. Bernal-Escobedo, and J. Santa, "Eco-efficient mobility in smart city scenarios," *Sustainability*, vol. 12, no. 20, pp. 1–15, 2020, doi: 10.3390/su12208443.

[57] I. Turgel, L. Bozhko, E. Ulyanova, and A. Khabdullin, "Implementation of the smart city technology for environmental protection management of cities: The experience of Russia and Kazakhstan," *Environ. Climate Technol.*, vol. 23, no. 2, pp. 148–165, Nov. 2019, doi: 10.2478/rtuect-2019-0061.

[58] X. Liu, Y. Song, K. Wu, J. Wang, D. Li, and Y. Long, "Understanding urban China with open data," *Cities*, vol. 47, pp. 53–61, Sep. 2015.

[59] Z. Chen, "Application of environmental ecological strategy in smart city space architecture planning," *Environ. Technol. Innov.*, vol. 23, Aug. 2021, Art. no. 101684, doi: 10.1016/j.eti.2021.101684.

[60] D. Andone, A. Ternauciuc, S. Vert, M. Mocofan, V. Mihaescu, D. Stoica, and R. Vasiu, "Learning with open cultural data–jecza museum for digiculture study case," in *Proc. Int. Conf. Adv. Learn. Technol. (ICALT)*, Tartu, Estonia, Jul. 2021, pp. 17–19.

[61] V. Roblek, I. Strugar, M. Mesko, M. P. Bach, and B. Jakovic, "E-democracy tools adoption: Experience of Austria, Croatia, Italy, and Slovenia," in *Proc. MIPRO*, Opatija, Croatia, Oct. 2020, pp. 1329–1335.

[62] J. Cano, R. Hernández, and S. Ros, "Distributed framework for electronic democracy in smart cities," *Computer*, vol. 47, no. 10, pp. 65–71, Oct. 2014, doi: 10.1109/MC.2014.280.

[63] *Recommendation CM/Rec(2009)1 and Explanatory Statement, La Democratie Électronique. Recommandation CM/Rec(2009)1 et exposé des Motifs*, E-Democracy, Strasbourg, France, 2009.

[64] P. Tang, "Telecom customer churn prediction model combining K-means and XGBoost algorithm," in *Proc. 5th ICMCCE*, 2020, pp. 1128–1131, doi: 10.1109/ICMCCE51767.2020.00248.

[65] M. Beshley, P. Veselý, A. Pryslupskyi, H. Beshley, M. Kyryk, V. Romanchuk, and I. Kahalo, "Customer-oriented quality of service management method for the future intent-based networking," *Appl. Sci.*, vol. 10, no. 22, pp. 1–38, 2020.

[66] E. Dizon and B. Pranggono, "Smart streetlights in smart city: A case study of Sheffield," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 4, pp. 2045–2060, Apr. 2022, doi: 10.1007/s12652-021-02970-y.

[67] L. Chen, D. Lu, M. Zhu, M. Muzammal, O. W. Samuel, G. Huang, W. Li, and H. Wu, "OMDP: An ontology-based model for diagnosis and treatment of diabetes patients in remote healthcare systems," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 5, May 2019, Art. no. 155014771984711, doi: 10.1177/1550147719847112.

[68] H. Liu, Y. Chuang, C. Liu, P. C. Yang, and C. Fuh, "Precise measurement of physical activities and high-impact motion: Feasibility of smart activity sensor system," *IEEE Sensors J.*, vol. 21, no. 1, pp. 568–580, Jan. 2021, doi: 10.1109/JSEN.2020.3015392.

[69] *WHO*. Accessed: Feb. 2023. [Online]. Available: https://www.who.int/fr/news-room/fact-sheets/detail/cancer

[70] S. Mainali, M. E. Darsie, and K. S. Smetana, "Machine learning in action: Stroke diagnosis and outcome prediction," *Frontiers Neurol.*, vol. 12, Dec. 2021, Art. no. 734345, doi: 10.3389/fneur.2021.734345.

[71] R. K. Pathan, F. I. Alam, S. Yasmin, Z. Y. Hamd, H. Aljuaid, M. U. Khandaker, and S. L. Lau, "Breast cancer classification by using multi-headed convolutional neural network modeling," *Healthcare*, vol. 10, no. 12, p. 2367, Nov. 2022, doi: 10.3390/healthcare10122367.

[72] T. Vaiyapuri, A. K. Dutta, I. S. H. Punithavathi, P. Duraipandy, S. S. Alotaibi, H. Alsolai, A. Mohamed, and H. Mahgoub, "Intelligent deep-learning-enabled decision-making medical system for pancreatic tumor classification on CT images," *Healthcare*, vol. 10, no. 4, p. 677, Apr. 2022, doi: 10.3390/healthcare10040677.

[73] H. Abbasimehr and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization," *Chaos, Solitons Fractals*, vol. 142, Jan. 2021, Art. no. 110511.

[74] R. Ma, X. Zheng, P. Wang, H. Liu, and C. Zhang, "The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method," *Sci. Rep.*, vol. 11, no. 1, p. 17421, Aug. 2021.

[75] H. Abbasimehr, R. Paki, and A. Bahrini, "A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting," *Neural Comput. Appl.*, vol. 34, no. 4, pp. 3135–3149, Feb. 2022.

[76] G. Li, K. Chen, and H. Yang, "A new hybrid prediction model of cumulative COVID-19 confirmed data," *Process Saf. Environ. Protection*, vol. 157, pp. 1–19, Jan. 2022.

[77] R. Tang, Z. Zhu, H. Yao, Y. Li, X. Sun, G. Hu, G. Xie, and Y. Li, "Integrating medical code descriptions and building text classification models for diagnostic decision support," in *Proc. IEEE 10th Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2022, pp. 612–613, doi: 10.1109/ichi54592.2022.00122.

[78] M. Akshay Kumaar, D. Samiayya, P. M. D. R. Vincent, K. Srinivasan, C.-Y. Chang, and H. Ganesh, "A hybrid framework for intrusion detection in healthcare systems using deep learning," *Frontiers Public Health*, vol. 9, Jan. 2022, Art. no. 824898, doi: 10.3389/fpubh.2021.824898.

[79] B. Walek and V. Fojtik, "A hybrid recommender system for recommending relevant movies using an expert system," *Exp. Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113452.

[80] J. Wang, P. Huang, H. Zhao, Z. Zhang, B. Zhao, and D. Lee, "Billion-scale commodity embedding for e-commerce recommendation in Alibaba," in *Proc. ACM SIGKDD*, 2018, pp. 839–848.

[81] A. Sanaullah, C. Yang, Y. Alexeev, K. Yoshii, and M. C. Herbordt, "Real-time data analysis for medical diagnosis using FPGA-accelerated neural networks," *BMC Bioinf.*, vol. 19, no. S18, Dec. 2018, doi: 10.1186/s12859-018-2505-7.

[82] P. R. Ovi, E. Dey, N. Roy, and A. Gangopadhyay, "ARIS: A real time edge computed accident risk inference system," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Aug. 2021, pp. 47–54, doi: 10.1109/SMARTCOMP52413.2021.00027.

[83] C. J. M. Casillas Mora, G. Ochoa Ruiz, and L. M. Aguilar Lobo, "IoT-based panel for real time traffic data monitoring in smart cities: A case study in the Guadalajara metropolitan zone," in *Proc. Int. Conf. Electron., Commun. Comput. (CONIELECOMP)*, Feb. 2019, pp. 1–12.

[84] M. Gao, K. Liu, and Z. Wu, "Personalisation in web computing and informatics: Theories, techniques, applications, and future research," *Inf. Syst. Frontiers*, vol. 12, no. 5, pp. 607–629, Nov. 2010.

[85] D. Poo, B. Chng, and J.-M. Goh, "A hybrid approach for user profiling," in *Proc. 36th Annu. Hawaii Int. Conf. Syst. Sci.*, 2003, pp. 1–14.

[86] F. M. Bianchi, A. Rizzi, A. Sadeghian, and C. Moiso, "Identifying user habits through data mining on call data records," *Eng. Appl. Artif. Intell.*, vol. 54, pp. 49–61, Sep. 2016.

[87] B. Ayesha, B. Jeewanthi, C. Chitraranjan, A. M. Perera, and A. S. Kumarage, "User localization based on call detail record," in *Intelligent Data Engineering and Automated Learning—IDEAL*, vol. 11871. Manchester, U.K.: Springer, 2019, pp. 411–423.

[88] A. Garcia-Davalos and J. Garcia-Duque, "User profile modelling based on mobile phone sensing and call logs," *Inf. Technol. Syst.*, vol. 1137, pp. 243–254, Jan. 2020.

[89] Y. Wei, M. Peng, and Y. Liu, "Intent-based networks for 6G: Insights and challenges," *Digit. Commun. Netw.*, vol. 6, no. 3, pp. 270–280, Aug. 2020, doi: 10.1016/j.dcan.2020.07.001.

[90] A. Farseev, L. Nie, M. Akbari, and T. Chua, "Harvesting multiple sources for user profile learning: A big data study," in *Proc. ICMR*, 2015, pp. 235–242, doi: 10.1145/2671188.2749381.

[91] X. Song, L. Nie, L. Zhang, M. Akbari, and T.-S. Chua, "Multiple social network learning and its application in volunteerism tendency prediction," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2015, pp. 213–222.

[92] C. Xie, H. Cai, Y. Yang, L. Jiang, and P. Yang, "User profiling in elderly healthcare services in China: Scalper detection," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 6, pp. 1796–1806, Nov. 2018, doi: 10.1109/JBHI.2018.2852495.

[93] M. Bohmer, B. Hecht, J. Schoning, A. Kruger, and G. Bauer, "Falling asleep with angry birds, Facebook and Kindle—A large scale study on mobile application usage," in *Proc. MobileHCI*, 2011, pp. 47–56, doi: 10.1145/2037373.2037383.

[94] B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery," in *Proc. 9th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2011, pp. 113–126.

[95] D. Ferreira, V. Kostakos, and A. K. Dey, "Aware: Mobile context instrumentation framework," *Frontiers ICT*, vol. 2, p. 6, Apr. 2015.

[96] A. J. Oliner, A. P. Iyer, I. Stoica, E. Lagerspetz, and S. Tarkoma, "Carat: Collaborative energy diagnosis for mobile devices," in *Proc. SenSys*, 2013, pp. 1–14.

[97] D. T. Wagner, A. Rice, and A. R. Beresford, "Device analyzer: Large-scale mobile data collection," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 4, pp. 53–56, Apr. 2014.

[98] A. S. Jacobs, R. J. Pfitscher, R. H. Ribeiro, R. A. Ferreira, L. Z. Granville, W. Willinger, and S. G. Rao, "Hey, Lumi! Using natural language for intent-based network management," in *Proc. USENIX Annu. Tech. Conf.*, Jul. 2021, pp. 1–16.

[99] N. Rajasinghe, J. Samarabandu, and X. Wang, "INSECS-DCS: A highly customizable network intrusion dataset creation framework," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Quebec, QC, Canada, May 2018, pp. 1–4.

[100] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Exp. Syst. Appl.*, vol. 41, no. 4, pp. 2065–2073, Mar. 2014.

[101] S. Arora. *Data Mining vs Machine Learning: The Key Difference*. Accessed: Feb. 2023. [Online]. Available: https://www.simplilearn.com/data-mining-vs-machine-learning-article

[102] *Data Mining vs Machine Learning: What is the Difference*. Accessed: Feb. 2023. [Online]. Available: https://www.epsi.fr/data-mining-machine-learning-difference/

[103] B. Srinu, P. N. L. Bhavana, B. T. Reddy, and B. Vaishnavi, "Machine learning based suicide prediction," in *Proc. ICCMC*, 2022, pp. 953–957, doi: 10.1109/ICCMC53470.2022.9754035.

[104] B. S. Atote, S. Zahoor, B. Dangra, and M. Bedekar, "Personalization in user profiling: Privacy and security issues," in *Proc. Int. IOTA*, 2016, pp. 415–417, doi: 10.1109/IOTA.2016.7562763.

[105] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu, "Multimodal graph-based reranking for web image search," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4649–4661, Nov. 2012.

[106] S. Fodeh, B. Punch, and P.-N. Tan, "On ontology-driven document clustering using core semantic features," *Knowl. Inf. Syst.*, vol. 28, no. 2, pp. 395–421, Aug. 2011.

[107] A. F. Y. Mohammed, S. M. Sultan, J. Lee, and S. Lim, "Deep-reinforcement-learning-based IoT sensor data cleaning framework for enhanced data analytics," *Sensors*, vol. 23, no. 4, p. 1791, Feb. 2023, doi: 10.3390/s23041791.

[108] S. P. Menon, P. K. Shukla, P. Sethi, A. Alasiry, M. Marzougui, M. T.-H. Alouane, and A. A. Khan, "An intelligent diabetic patient tracking system based on machine learning for E-health applications," *Sensors*, vol. 23, no. 6, p. 3004, Mar. 2023, doi: 10.3390/s23063004.

[109] A. Amato and V. Di Lecce, "Data analysis for information discovery," *Appl. Sci.*, vol. 13, no. 6, p. 3481, Mar. 2023, doi: 10.3390/app13063481.

[110] Y. Wu, J. Ma, X. Huang, S. Ling, and S. Weidong Su, "DeepMMSA: A novel multimodal deep learning method for non-small cell lung cancer survival analysis," in *Proc. IEEE Int. Conf. SMC*, Oct. 2021, pp. 1468–1472, doi: 10.1109/SMC52423.2021.9658891.

[111] F. He, J. Zhou, Z.-K. Feng, G. Liu, and Y. Yang, "A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm," *Appl. Energy*, vol. 237, pp. 103–116, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.055.

[112] R. Wang, C. Li, W. Fu, and G. Tang, "Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3814–3827, Oct. 2020, doi: 10.1109/TNNLS.2019.2946414.

[113] *The Role of Natural Language Processing in AI*. Accessed: Dec. 2022. [Online]. Available: https://online.york.ac.uk/the-role-of-natural-language-processing-in-ai/

[114] B. Trstenjak, S. Mikac, and D. Donko, "KNN with TF-IDF based framework for text categorization," *Proc. Eng.*, vol. 69, pp. 1356–1364, 2014, doi: 10.1016/j.proeng.2014.03.129.

[115] N. Guarino, S. Uniti, and P. Giaretta, "Ontologies and knowledge bases. Toward a terminological clarification," *Towards Very Large Knowl. Bases*, vol. 25, no. 32, pp. 307–317, 1995.

[116] T. Wei, Y. Lu, H. Chang, Q. Zhou, and X. Bao, "A semantic approach for text clustering using WordNet and lexical chains," *Exp. Syst. Appl.*, vol. 42, no. 4, pp. 2264–2275, Mar. 2015.

[117] V. Moustaka, A. Vakali, and L. G. Anthopoulos, "Citydna: Smart city dimensions' correlations for identifying urban profile," in *Proc. Int. Conf. WWW*, 2017, pp. 1167–1172, doi: 10.1145/3041021.3054714.

[118] Y. Ma, J. Mao, Z. Ba, and G. Li, "Location recommendation by combining geographical, categorical, and social preferences with location popularity," *Inf. Process. Manage.*, vol. 57, no. 4, Jul. 2020, Art. no. 102251, doi: 10.1016/j.ipm.2020.102251.

[119] M. Ye, P. Yin, W. Lee, and D. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. Int. Conf. SIGIR*, 2011, pp. 325–334, doi: 10.1145/2009916.2009962.

[120] S. Xiang, L. Yuan, W. Fan, Y. Wang, P. M. Thompson, and J. Ye, "Multi-source learning with block-wise missing data for Alzheimer's disease prediction," in *Proc. Int. Conf. KDD*, 2013, pp. 185–193, doi: 10.1145/2487575.2487594.

[121] J. Zhang and J. Huan, "Inductive multi-task learning with multiple view data," in *Proc. Int. Conf. KDD*, 2012, pp. 543–551.

[122] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," 2020, *arXiv:2001.05566*.

[123] A. Andreas, C. X. Mavromoustakis, H. Song, and J. M. Batalla, "CNN-based emotional stress classification using smart learning dataset," in *Proc. IEEE Int. Conf. iThings GreenCom CPSCom SmartData Cybermatics*, Espoo, Finland, Aug. 2022, pp. 549–554.

[124] P. Zhao, S. C. H. Hoi, J. Wang, and B. Li, "Online transfer learning," *Artif. Intell.*, vol. 216, pp. 76–102, Nov. 2014.

[125] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[126] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, Oct. 2018.

[127] X. Chen, M. Hou, T. Tang, A. Kaur, and F. Xia, "Digital twin mobility profiling: A spatio-temporal graph learning approach; digital twin mobility profiling: A spatio-temporal graph learning approach," in *Proc. Int. Conf. HPCC/ DSS/SmartCity/DependSys*, Haikou, Hainan, China, Dec. 2021, pp. 20–22.

[128] M. Liu, L. Li, Q. Li, Y. Bai, and C. Hu, "Pedestrian flow prediction in open public places using graph convolutional network," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 7, p. 455, Jul. 2021.

[129] E. Carter, P. Adam, D. Tsakis, S. Shaw, R. Watson, and P. Ryan, "Enhancing pedestrian mobility in smart cities using big data," *J. Manage. Analytics*, vol. 7, no. 2, pp. 173–188, Apr. 2020.
[130] F. Harrou, A. Dairi, A. Zeroual, and Y. Sun, "Forecasting of bicycle and pedestrian traffic using flexible and efficient hybrid deep learning approach," *Appl. Sci.*, vol. 12, no. 9, p. 4482, Apr. 2022.
[131] L. Cornejo-Bueno, E. C. Garrido-Merchán, D. Hernández-Lobato, and S. Salcedo-Sanz, "Bayesian optimization of a hybrid system for robust ocean wave features prediction," *Neurocomputing*, vol. 275, pp. 818–828, Jan. 2018, doi: 10.1016/j.neucom.2017.09.025.

KARIDJA DOMINIQUE CHRISTELLE ADJE received the M.Eng. degree in network and telecommunication from the African Higher School of ICT (ESATIC), Côte d'Ivoire, in partnership with the Higher School of Communication of Tunis (SUP'COM). She is currently pursuing the dual Ph.D. degree in cotutelle with the Mediatron Research Laboratory, SUP'COM, University of Carthage, Tunisia, and with the LIMOS, University of Clermont Auvergne, France. Her research interests include open data, smart city, data processing, machine learning, and security.

ASMA BEN LETAIFA received the degree in telecom engineering from the Higher School of Communications (SUP'COM), University of Carthage, Tunisia, and the joint Ph.D. degree from SUP'COM, University of Carthage, and from Université de Bretagne Occidentale (UBO), Brest, France. She is currently an Assistant Professor and a member of the Mediatron Research Laboratory, SUP'COM, University of Carthage. She is the author of several courses on telecommunications services, network modeling with queuing theory, web content, cloud architectures and virtualization, massive big data content, and machine learning algorithms. She is the coauthor of the "Linux practices" MOOC on the FUN platform. Her research interests include telecom services, cloud and mobile cloud architectures, service orchestration, big data, and quality of experience in an SDN/NFV environment. She is the author or coauthor of several articles on these subjects.

MAJED HADDAD received the Diploma degree in electrical engineering from the National Engineering School of Tunis, Tunisia, in 2004, the master's degree from the University of Nice Sophia Antipolis, France, in 2005, and the Ph.D. degree in electrical engineering from the Eurecom Institute, in 2008. In 2009, he joined the France Telecom Research and Development, as a Postdoctoral Research Fellow. In 2011, he joined Avignon University, France, as a Researcher Assistant. From 2012 to 2014, he was a Research Engineer with INRIA Sophia-Antipolis, France, under the INRIA Alcatel-Lucent Bell Laboratory Fellowship. He has been an Assistant Professor with Avignon University, since 2014. He has published more than 50 research papers in international conferences, journals, book chapters, and patents. His research interests include radio resource management, heterogeneous networks, green networks, complex networks, and game theory. He is the TPC chair, a TPC member, and a reviewer of various prestigious conferences and journals.

OUSSAMA HABACHI received the engineering degree in computer science from the National School of Computer Sciences (ENSI), in September 2008, the M.Sc. degree in network and communications from the University of Pierre and Marie Curie (UPMC), Paris, France, in 2009, and the Ph.D. degree in computer science from the University of Avignon, in September 2012. In October 2012, he joined INRIA Sophia-Antipolis, as a Postdoctoral Research Fellow, working on optimal impulsive control systems, cognitive radio, and game theory. From 2014 to 2021, he was an Assistant Professor with the University of Limoges. He is currently a Full Professor with the University of Clermont Auvergne and a member of the LIMOS.