Smart City Big Data Analytics

An Advanced Review

Kamran Soomro [*], Nasir Bhutta [†], Zaheer Khan[‡], Muhammad Atif Tahir [§]

Article Type:

Advanced Review

Abstract

With the increasing role of ICT in enabling and supporting smart cities, the demand for big data analytics solutions is increasing. Various artificial intelligence, data mining, machine learning and statistical analysis based solutions have been successfully applied in thematic domains like climate science, energy management, transport, air quality management and weather pattern analysis. In this paper we present a systematic review of the literature on smart city big data analytics. We have searched a number of different repositories using specific keywords and followed a structured data mining methodology for selecting material for the review. We have also performed a technological and thematic analysis of the shortlisted literature, identified various data mining / machine learning techniques and presented the results. Based on our analysis we also present a classification model that studies four aspects of research in this domain. These include data models, computing models, security and privacy aspects and major market drivers in the smart cities domain. Moreover, we present a gap analysis of and identify future directions for research. For our thematic analysis we identified the themes smart city governance, economy, environment, transport and energy. We present the major challenges in these themes, the major research work done in the field of data analytics to address these challenges and future research directions.

[*]Computer Science Research Centre, UWE, Bristol, UK
[†]King Faisal University, Al-Ahsa, Saudi Arabia
[‡]Computer Science Research Centre, UWE, Bristol, UK
[§]National University of Computer and Emerging Sciences, Karachi Campus, Pakistan

INTRODUCTION

Various statistical, machine learning and data mining techniques have proved to be revolutionary with respect to Smart City Data Analytics. Deep learning models are one of the most cutting-edge technologies that are introduced for various applications for smart city services (Mohammadi, Al-Fuqaha, Sorour, & Guizani, 2018; Mohammadi, Al-Fuqaha, & Oh, 2018; Zaouali, Rekik, & Bouallegue, 2018). Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM), Autoencoders (AEs), Generative Adversarial Networks (GANs), and Deep Belief Network (DBNs) are widely known Deep Learning models that are currently being investigated for various Smart City Applications including Energy consumption prediction, behaviour detection, security threat identification, traffic sign detection, face recognition and human activity recognition (Mohammadi et al., 2018). Similarly learning from web data is crucial for many applications in smart cities and several deep learning approaches have been proposed recently based on deep probabilistic framework (Niu, Tang, Veeraraghavan, & Sabharwal, 2018) to learn from web data.

According to the Deloitte Consulting Group (van Dijk & Teuben, 2015):

> "A city is smart when investments in (i) human and social capital, (ii) traditional infrastructure and (iii) disruptive technologies fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance."

Therefore, the key elements of a smart city, according to Deloitte, are:

- its human and social capital,
- its traditional infrastructure, and
- disruptive technologies such as Artificial Intelligence, machine learning, data mining, augmented reality, distributed ledger technologies, IoTs, etc

Investment in the nexus of these three elements leads to certain benefits:

- sustainable economic growth,
- high quality of life, and

- judicious use and management of natural resources resulting in a reduced ecological footprint.

- prediction of future scenarios for sustainable cities

The nature of these key elements and benefits and their complex inter-relationships is such that smart solutions are multi-sectoral by necessity. These span domains such as smart mobility, smart safety, smart energy, smart water and waste, smart health, smart education and smart government etc. One of the major transformations that a smart city brings is in the sharing of data and information between these various domains. Traditionally these areas have acted as information silos, with policy-making in each being restricted to the confines of the respective area. In smart cities these barriers are expected to be broken, and information flows freely between them with policy-making being an integrated and holistic process. The resultant ICT solutions are thus multi-faceted and complex by nature and heavily rely on AI, data mining, statistical and machine learning methods to generate new information and knowledge from multi-disciplinary data sets. A more detailed list of the goals, challenges, and smart domains is shown in Table 1. This multi-faceted and challenging nature of smart cities provide potent opportunities for the data mining and knowledge management communities to test existing models and algorithms as well as design new ones.

Table 1: Smart city goals, challenges and domains (van Dijk & Teuben, 2015).

Goals	Challenges	Domains
Economic growth	• Controlled transition of the labour market due to automation	• Smart Mobility
Quality of life	• Winning the war on talent between metropolitan areas	• Smart Safety

| Reduced ecological footprint | • Social cohesion, inclusiveness, solidarity
• Secure digital environment, privacy
• Resilience | • Smart Energy, Water & Waste
• Smart Buildings & Living
• Smart Health
• Smart Education
• Smart Finance
• Smart Tourism & Leisure
• Smart Retail & Logistics
• Smart Manufacturing & Construction
• Smart Government |

Table 2 shows the various disruptive technologies in the context of smart cities. These disruptive technologies are causing the very nature of business-as-usual in cities to transform by informing evidence-based decision-making (MGI, 2013). For example, by providing new methods and paradigms for communication, Internet of Everything is transforming domains from transport and energy management to waste management (enerquire, 2017; Treadwell, 2017). Blockchain, on the other hand, is disrupting how global economies function and do business securely and transparently. However, since in this paper we are focusing on big data analytics in smart cities, we will cover only that in subsequent sections.

Table 2: Disruptive technologies in the context of smart cities (van Dijk & Teuben, 2015).

Data Mining / Artificial intelligence	Co-creation
Internet of Everything	Drones
Social robotics	Renewable energy
Gamification	3D printing
Sharing economy	Crowd sourcing
Social media / Digital platforms	Blockchain
Big data	Mobile
Cloud	Self-organisation

In this paper we present an overview of how big data analytics can help enable smart cities. Our specific focus is on smart city challenges that can be addressed by big data analytics and in this regard we also review past and future trends. This will provide a

holistic multi-disciplinary view of challenges faced by the smart cities communities where the data mining and knowledge management communities can contribute. The rest of this paper is organised as follows. We begin by introducing our research methodology, including inclusion and exclusion criteria, search engines, data mining techniques, etc. We then present an introduction of big data technologies followed by how big data technologies are used in smart cities. This includes a description of the tools and techniques used as well as a description of various thematic areas in smart cities. We then conclude our review and draw some conclusions.

RESEARCH METHODOLOGY

It is important for a structured literature review to follow a consistent, reproducible methodology. There is a wealth of literature available on subjects such as data analytics, smart cities, big data and big data analytics. As this review focuses on smart city data analytics, we also restricted our search to this subject. We also adopted a data mining-based approach for sifting through and analysing the literature (Haneem, Kama, Ali, & Selamat, 2017). We queried a number of repositories such as ACM, arXiv, DBLP, Google Scholar, IEEExplore, Inspire, JSTOR, ScienceDirect and Scopus. Search terms included *smart city big data analytics* and *smart city analytics*. We searched these repositories using the terms provided, and initially selected 100 most recent papers based on their titles and abstracts. We further filtered these papers by going through their abstracts, introductions and conclusions and short listed about 66 papers. We then created a wordcloud based on the abstracts of these 66 selected papers, shown in Figure 1. Since words like *smart*, *analytics*, *big*, *data* and *cities* were part of our search terms, it is no surprise that they occur quite frequently in the abstracts of the filtered papers. Therefore, we can safely ignore them for the purposes of our analysis. Of the remaining frequently-occurring words, *economic*, *governance*, *energy* etc are significant. We arranged such significant words into a taxonomy by folding them into common themes, shown in Figure 2. The leaf nodes depict terms identified by the wordcloud while the parent nodes depict the common themes we have identified in our analysis. For example, the words *congestion*, *vehicle*, and *traffic* all fall under the common theme of *Transport*. We also chose

Figure 1: Most frequently used words mined from selected abstracts.

to study different aspects of those themes by using the other frequent keywords like *data* and *computing*.

Our analysis confirms what were suggested by Vinod Kumar (2015) and Chatterjee and Kar (2015) as the six most important themes of smart cities; smart economy, smart governance, smart mobility, smart environment, smart living and smart people. This work is different from Chauhan, Agarwal, and Kar (2016) in the way that we mainly focus on the current challenges present on the above mentioned five themes and contributions of big data analytics to solve them. Existing work focuses more on the big data characteristics of various themes. The specific aspects were also chosen on the basis of their relevance to big data analytics, e.g. both the data and computing models directly influence how the analytics are applied and how effective they are. Privacy is an important concern in any contemporary software system and the recent focus on privacy legislation has made it further important. Similarly market drivers can act as an indication of the widespread adoption of the various technologies and allow us to ascertain their popularity.

In the second phase we searched the above repository again with keywords specific to the
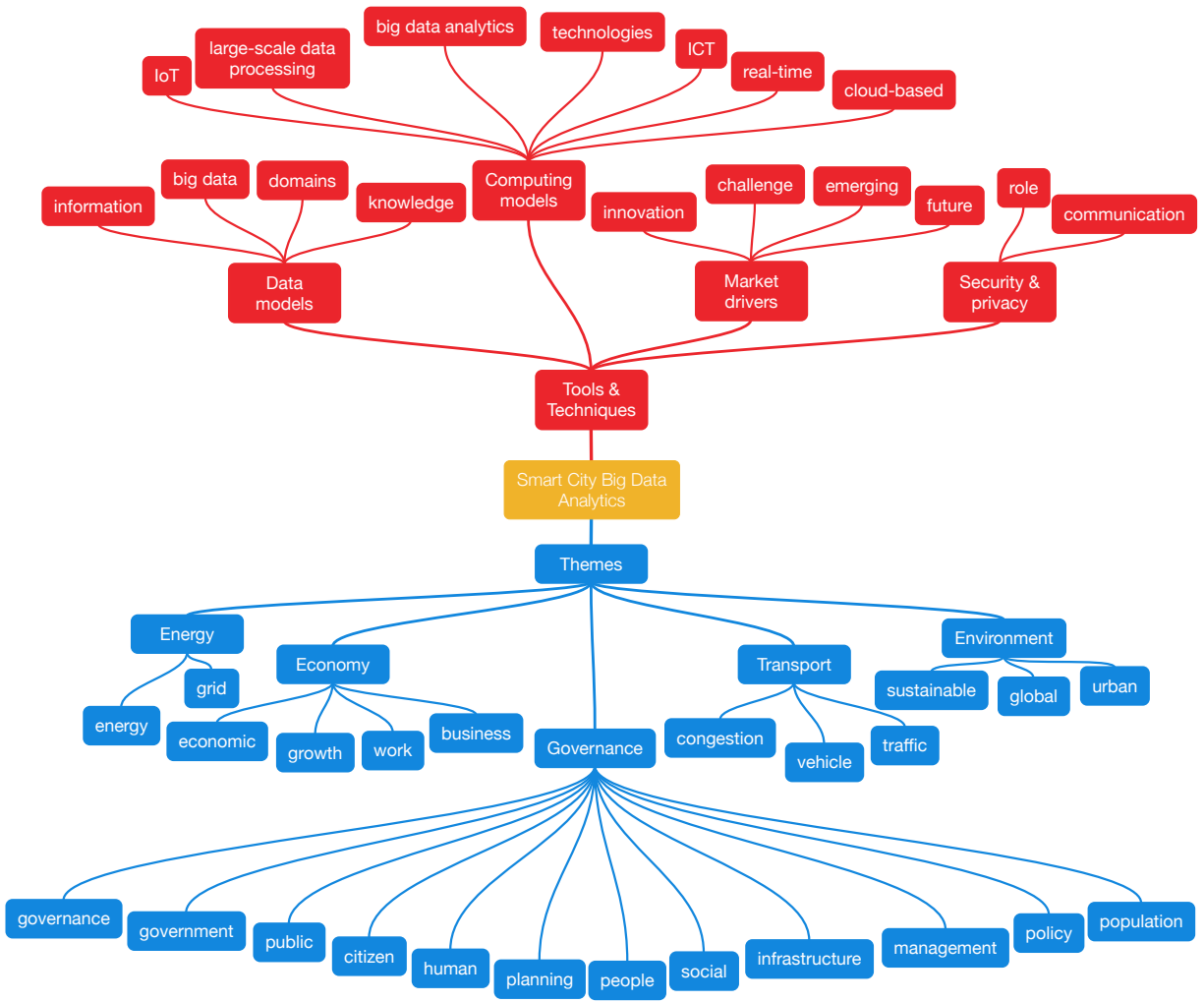
Figure 2: Classification model of various themes in smart cities.

identified themes, e.g. *city governance data analytics* and *city governance and data analysis* for the governance theme, *Smart City Transportation* and *Smart City Transportation and Big Data Analytics* for the transport theme and again short listed a number of papers based on their titles and abstracts. Table 3 shows the search terms and repositories queried for each repository for each theme. After this we critically reviewed the papers for our analysis. We then studied the literature from a number of aspects, e.g. what are the key trends currently in vogue in smart cities data analytics? Who are the major drivers (commercial, public, government, etc) behind these trends? What is the major research done in this area within the past 5 to 10 years? What are the future challenges and directions? To identify the latest trends we also reviewed different reports by various organisations on smart cities. The results are chronicled in the following sections.

Table 3: Keywords and repositories search repositories for various themes.

Theme	Keywords	Repositories
Governance	- city governance data analytics - city governance and data analysis, security and ICT - smart city governance data analytics - smart city data analytics - urban governance, smart city, data analytics	- Google scholar and Google search - IEEE Xplore - Science Direct
Environment	- smart city environments - environmental analytics - smart environments - smart environment big data - air quality big data - climate big data	- Google Scholar - Science Direct - IEEExplore - JStore - arXiv - DBLP - SpringerLink

Theme	Keywords	Repositories
Transport	- Smart City Transportation - Smart City Transportation and Big Data Analytics - Smart City Transportation Big Data Analytics - Smart City Big Data Analytics, Urban, Transportation - Smart City Transportation Security - Smart City Mobility and Big Data Analytics - Smart City Mobility Security	- Google Search Engine, - Google Scholar, - IEEExplore, - Science Direct
Economy	- Economic Forecasting - Machine Learning - Economics and Data Science - Economy and Big Data - Deep Learning and Economy - Smart Economy	- Google Scholar - Science (Journal) - IEEExplore - Springer - Science Direct
Energy	- Energy Big Data Analytics - Smart City Energy Big Data Analytics - Smart City Energy Security	- Google Search Engine - Google Scholar - IEEExplore - Science Direct

BIG DATA ANALYTICS

"Big data analytics is the process of examining large and varied data sets – i.e., big data – to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more-informed business decisions" (Rouse, 2017). In recent years big data has emerged as an important disruptive technology that has revolutionised how businesses operate. Recently Visa used big data analytics to reduce the time for processing 73 billion transactions from one month to 13 minutes (John Walker, 2014). Similarly big data analytics has brought about significant impact in many other domains such as climate change analysis, weather pattern analysis, real estate markets etc (Barkham, Bokhari, & Saiz, 2018; H. & Vipin, 2014; Ismail, Majid, Zain, & Bakar, 2016).

Big data is characterised by the four Vs; Volume, Variety, Velocity and Veracity (Mohanty, 2015). Given these characteristics processing and analysing big data in a reasonable amount of time is generally beyond the capabilities of a single computer system. This is why big data analytics necessitates the use of distributed processing techniques. In recent years several paradigms have emerged that harness the power of distributed computing. These include cluster computing (Ganelin, Orhian, Sasaki, & York, 2016) and cloud computing (Khan, Anjum, Soomro, & Tahir, 2015) as well as fog and edge computing (Simmhan, 2018). These paradigms are proving to be instrumental in the success of big data analytics. We will cover different examples of these in subsequent sections.

From a more technical point-of-view, big data analytics consists of a number of techniques that are used to analyse the data. These include, but are not limited to association rule learning, classification tree analysis, genetic algorithms, machine learning, regression analysis, sentiment analysis, social network analysis and deep learning (Stephenson, 2013). As with many other domains, big data analytics has also been applied in smart cities.

BIG DATA ANALYTICS IN SMART CITIES

In the past few years big data analytics has emerged as an important disruptive technology in enabling the "smartness" of smart cities (Bassoo, Ramnarain-Seetohul, Hurbungs, Fowdur, & Beeharry, 2018). Other disruptive technologies such as Internet of Things (IoT) generate big data that needs to be processed and analysed in short amounts of time to enable smart city governments to react to emerging patterns and trends in an efficient manner. The importance of big data analytics using AI, data mining, machine learning or statistical methods in smart cities cannot be stressed enough. Its benefits cut across any number of domains such as transportation planning, urban planning, smart buildings, weather prediction and analysis etc. We conducted an extensive literature review of the various ways in which big data analytics are being applied to smart cities. Thematic and technological meta-analyses of this literature revealed a number of cross-cutting concerns that are common across the various themes we identified. We subsequently developed a classification model from these concerns, shown in Figure 2.

In the field of smart cities big data analytics is the key technology to enable evidence-based decision making, developing and providing new public services as well as allowing citizens to co-create new public services (Dustdar, Nastic, & Scekic, 2016b; Lv et al., 2018, etc). Big data can also be used to monitor government initiatives (K. R. Malik, Sam, Hussain, & Abuarqoub, 2018) as well as provision of various user-related public services (Zhang, Zhang, Zhang, Shi, & Zhong, 2016). Monitoring and predicting air quality is another example of how big data can be used in smart cities on which there is a wealth of literature (GSMA, 2018; Zheng et al., 2015, etc.). Traditional Machine Learning approaches such as Support Vector Machines, Decision Tree, Nearest Neighbour classifier are still widely used approaches for the analysis of big data. For large datasets, Projection-SVM is a recent distributed implementation of kernel support vector using subspace partitioning (Singh & Mohan, 2018). Nguyen, Le, Nguyen, Phung, and Webb (2018) have investigated Bayesian kernel machines using Stein Variational Gradient Descent for Big Data analysis resulting in efficient and fast learning method. Bayesian approach is quite popular for some applications of smart city services. For example, recently a Bayesian approach to residential property valuation based on built environment and house characteristics is proposed by Liu et al. (2018). The proposed model is able to predict the main feature of submarket in house attribute similarity, location proximity, and substitutability. In another study, SVM is also used by Preda, Oprea, Bâra, and Belciu (Velicanu) (2018) for the forecasting of photovoltaic (PV) output for efficient renewable energy systems in Big Data Analytics Context. Support Vector Machine (SVM) and linear regression are investigated as machine learning approaches in this study and significant improvement in performance while introducing machine learning process. Iyengar, Lee, Irwin, Shenoy, and Weil (2018) use Bayesian inference which is a data-driven approach to identify the least energy efficient buildings in a city with large population. Bayesian inference is being used to estimate various parameters distribution of a building. A case study from a city containing more than 10,000 buildings show that more than half of the buildings required energy improvement measures.

Mohammadi and Al-Fuqaha (2018) have argued that most of the data generated by big data is unlabelled. Thus, semi-supervised learning is important for the proper utilization of unlabelled data. A three-level learning framework is proposed that matches the hierarchical

description of big data generated by smart cities.

In this section we first present the various tools and techniques that are used for different aspects in smart city projects. These are followed by the results of the thematic analysis.

Tools and Techniques

A number of different approaches, tools and techniques have been used to perform big data analytics in smart cities. We look at these from four different perspectives:

- tools and techniques used for data modelling, storage, transfer and management,
- tools and techniques used for computing, i.e. processing and analysing the data,
- tools and techniques used for managing security and privacy, e.g. authentication and authorisation, encryption, anonymisation etc, and
- major market drivers such as commercial vendors, research community, public sector, etc. that are supporting the discussed approaches.

Each of these is discussed subsequently.

Data Models

Relational databases are still the most commonly used storage mechanism for smart city applications Eirinaki et al. (2018); Motta, You, Sacco, and Ma (2014); Peters-Anders, Khan, Loibl, Augustin, and Breinbauer (2017). A number of tools and techniques are used to store, transfer and manage big data in smart cities. This is especially challenging because of the nature of big data discussed previously. Since it is generally large in volume, the databases used to store it must necessarily be distributed to be able to leverage multiple data stores. For this reason data storage solutions such as HDFS (Anisetti et al., 2018; Hu et al., 2018; Pei, Li, & Tong, 2018) and Amazon EC2 (Lu et al., 2011) are commonly used. Due to its frequent updates (velocity), big data is also, often by nature, time series data, and thus time series databases such as InfluxDB are also found in literature (Tahat, Aburub, Al-Zyoude, & Talhi, 2018). As big data is also often from many sources and thus, in a variety of formats, traditional relational databases such as MySQL, Oracle etc. that impose strict structural restrictions on the data are unsuitable. This paves way for more unstructured databases

such as MongoDB (Pei et al., 2018; Tahat et al., 2018).

Computing Models

As with data models, the nature of big data also necessitates different computing models. Moreover, since smart cities are designed to provide public information services to citizens, this also necessitates unique computing models. Cloud, fog and edge computing are important and emerging technologies in this regard (Bibri, 2017). On the backend distributed data processing technologies such as Hadoop (Anisetti et al., 2018; Ghaemi, Alimohammadi, & Farnaghi, 2018; Hu et al., 2018; Lu et al., 2011; Pei et al., 2018; Wang & Eick, 2017) and Microsoft Azure are popular (Tahat et al., 2018).

Our survey also found that machine learning is by far the most common method of performing big data analytics, e.g. Anisetti et al. (2018); Ghaemi et al. (2018); Le (2018); Lu et al. (2011); Pei et al. (2018) just to name a few. Machine learning techniques such as Artificial Neural Networks (ANNs), clustering and SVMs have been shown to be very successful for performing the sorts of analytics that are required (Le, 2018; Pei et al., 2018; Tahat et al., 2018; Zheng, Liu, & Hsieh, 2013).

Security and Privacy

There is a lot of work in smart city security and mostly standard security and privacy approaches have been applied (Khan, Pervez, & Abbasi, 2017). This is an important area of smart cities is made even more important due to the recent spotlight on security and privacy in the IT domain overall. The introduction of General Data Protection Regulation (GDPR) is one result of such a focus. As a result, research in this area is ramping up. Anisetti et al. (2018) propose a system that uses model-based definitions of public policies to ensure compliance with the GDPR.

Market Drivers

Smart city technologies are being driven by major players such as IBM (Wedgwood & Howard, 2015), GSMA (GSMA, 2018), Deloitte (van Dijk & Teuben, 2015) and many more

industrial and private players. Moreover, many open source projects have also emerged that address the challenges of smart cities.

Thematic Areas

Governance

Smart city governance (Bolívar, 2015; Meijer & Bolívar, 2016) is a complex phenomenon. It requires stakeholder engagement for increased public-private partnerships; citizen participation in planning for co-designing cities; data collection from built-environment, public agencies such as health, mobility, environment, security and socio-economic activities for analysis and decision making. In this respect, smart governance uses ICT tools for data collection, processing and analysis in order to better understand the state of city, improve communication, enhance public service provisioning and promote transparency by introducing evidence-based policies (Javed, Khan, & McClatchey, 2018; Khan, Dambruch, et al., 2017). Since the focus on this survey paper is data analysis, we will only cover data and technology related aspects.

The PwC report highlights the importance of Data Driven Cities (PWC, n.d.) for variety of smart city applications across the world and recommends technology development and introduction strategy for long-term sustainability of data based cities; horizontally (i.e. across departments) and vertically (i.e. thematic and level of governance) integrated information platform for consistent and compatible data processing; and, measurable indicators of technology in city management such as financial gains, efficiency in governance processes, etc. Similarly, Barns (2017) discusses role of urban data platforms or dashboards in supporting the city governance. She argues that data platforms contribute as a key element for the development of new governance models for smart cities; enabling various city stakeholders to test and discuss potentials and pitfalls of data-driven initiatives in smart governance.

This part of the survey attempts to review literature, specifically, focusing on urban governance and data analytics. For this section, about 160 relevant articles were shortlisted based on their title and keywords. Most of these shortlisted papers either focused on general data analytics in smart cities or talked about a specific theme such as energy or mobility

or health or emergency response or urban planning, etc. After reviewing abstracts of these articles, about 20 papers were considered suitable for governance theme review.

The above searched literature indicates that in the past few years there has been considerable increase in literature about smart city data analytics. Most noticeable articles which appeared in the above search criteria and highlighted smart city data, technology (e.g. IoT, sensors, cloud, fog, edge, GIS, AI, Big Data, data mining, machine learning, semantics, Human-Computer-Interaction, privacy and security) and system architectures applicable to variety of city applications are (Al Nuaimi, Al Neyadi, Mohamed, & Al-Jaroodi, 2015; Barns, 2017; Bawany & Shamsi, 2015; Bibri, 2017; Bibri & Krogstie, 2017; Bilal, Khalid, Erbad, & Khan, 2018; Chen et al., 2015; El Mendili, El Idrissi, & Hmina, 2016; Faieq, Saidi, Elghazi, & Rahmani, 2017; Habibzadeh et al., 2018; Hashem et al., 2016; Khan et al., 2015; Osman, Elragal, & Bergvall-Kåreborn, 2017; PWC, n.d.; Saggi & Jain, 2018; Silva, Khan, & Han, 2018; Susmitha & Jayaprada, 2017; Yin et al., 2015; Zotano & Bersini, 2017).

According to literature review, ICT is an essential component in smart city governance to gain necessary intelligence that provides basis for evidence-based decision-making. At the core of ICT based solutions are three main components: i) data collection through sensors, Internet-of-Things, smart phones, remote sensing (e.g. satellite or in-situ) and city databases; ii) data processing and/or pre-processing (e.g. filtering, data quality and format translations); iii) data analysis using machine learning, data mining and other statistical algorithms to generate new knowledge in various cross-thematic applications such as mobility (Docherty, Marsden, & Anable, 2017; Peters-Anders et al., 2017; Rathore et al., 2017), energy (Antonić, Marjanović, Pripužić, & Žarko, 2016; Carli, Albino, Dotoli, Mummolo, & Savino, 2015; Silva, Khan, & Han, 2017), health (Anisetti et al., 2018; Farahani et al., 2018), environment (Antonić et al., 2016), public services (Pérez-González & Díaz-Díaz, 2015; Zhang et al., 2016), economy (Chatfield & Reddick, 2017; Saggi & Jain, 2018; Zaman et al., 2017), waste management (Digiesi, Facchini, Mossa, Mummolo, & Verriello, 2015), social analysis (Kousiouris et al., 2018; Terroso-Saenz, Gonzalez-Vidal, Cuenca-Jara, & Skarmeta, 2017), waste water management (Edmondson et al., 2018), urban planning (Eirinaki et al., 2018; Pettit et al., 2018; Rathore, Ahmad, Paul, & Rho, 2016a), tourism and cultural heritage (Sun, Song, Jara, & Bie, 2016), buildings (Linder, Vionnet, Bacher, & Hennebert,

2017), agriculture (Kamilaris, Gao, Prenafeta-Boldu, & Ali, 2016), emergency response (Abu-Elkheir, Hassanein, & Oteafy, 2016), etc. The above three components - with some variations - are common among the most of smart city data analytics literature e.g., Khan et al. (2015); Rathore et al. (2016a); Zhang et al. (2016). The following Table 4 presents coverage of topics in the selected literature.

Table 4: Survey summary.

Topics covered	Survey articles
City governance and challenges	Aguilera et al. (2017); Anisetti et al. (2018); Babar and Arif (2017); Carli et al. (2015); Chatfield and Reddick (2017); Dustdar et al. (2016a); Eirinaki et al. (2018); Gil-Garcia et al. (2016); Khan et al. (2015); Khan, Dambruch, et al. (2017); Kourtit and Nijkamp (2018); Lv et al. (2018); Meijer and Bolívar (2016); Motta et al. (2014); Pan et al. (2016a); Pérez-González and Díaz-Díaz (2015); Rathore et al. (2016b); Zhang et al. (2016)
Urban and Citizen Data	Aguilera et al. (2017); Anisetti et al. (2018); Babar and Arif (2017); Barns (2017); Carli et al. (2015); Eirinaki et al. (2018); Khan et al. (2015); Khan, Dambruch, et al. (2017); Kourtit and Nijkamp (2018); Lv et al. (2018); K. R. Malik et al. (2018); Motta et al. (2014); Pan et al. (2016a); Rathore et al. (2016b); Sun et al. (2016); Zhang et al. (2016); Zotano and Bersini (2017)
Technology, system architecture or infrastructure	Aguilera et al. (2017); Anisetti et al. (2018); Babar and Arif (2017); Barns (2017); Carli et al. (2015); Dustdar et al. (2016a); Eirinaki et al. (2018); Javed et al. (2018); Khan et al. (2015); Khan, Dambruch, et al. (2017); Lv et al. (2018); Motta et al. (2014); Pan et al. (2016a); Rathore et al. (2016b); Sun et al. (2016); Zhang et al. (2016); Zotano and Bersini (2017)
Security and privacy	Anisetti et al. (2018); Dustdar et al. (2016a); Khan et al. (2015)

The transformation from traditional governance to ICT enabled smart governance (Bolívar, 2015; Meijer & Bolívar, 2016) is gaining momentum where open data or other data platforms are providing useful contributions (Barns, 2017; Pérez-González & Díaz-Díaz, 2015). On the one hand increasing Big urban data provide greater opportunities for new forms of city governance; on the other hand there is no systematic and coherent data system to manage unstructured, semi-structured and structured heterogeneous urban data in order to perform data analysis (Kourtit & Nijkamp, 2018). Barns (2017) and Kourtit and Nijkamp (2018) discuss the role of urban data platforms or dashboard in supporting the city governance. These data platforms contribute as key elements for the development of new governance models for smart cities and forums in which various city stakeholders can test and

16

discuss potentials and pitfalls of data-driven initiatives in smart governance (Barns, 2017). Similarly, (Gil-Garcia et al., 2016) presents different dimensions (e.g. technology, citizen engagement, openness, evidence-based decision making, creativity & innovation, sustainability, integration, effectiveness and efficiency) of the word 'smartness' in understanding and developing smart city governments. We agree with authors that technological advances should be equally met by advances in government management and policy for smart governance. This is also supported by Meijer and Bolívar (2016) who argue that smart city governance is not just a technological issue rather it is a complex process of institutional change with more appealing socio-technical governance focusing on both economic gains and other public values. Others refer this approach as 'collective smart' which is necessary value-driven context of cyber human smart cities (Dustdar et al., 2016b). Likewise Kourtit and Nijkamp (2018) argue that smart city management will fail without human or cognitive capital and lack of innovativeness and creativity. Hence cities of future needs both technical and socio-economic dimensions to be considered for urban policy making.

Among more technology savvy literature, (Lv et al., 2018) proposes a 3D GIS and cloud-based government affairs service platform that enables various city stakeholders to use city data (acquisition and management), analyse (Hadoop/map-reduce, data mining and statistical) and create new services developed through a multi-tier cloud-based architecture. Dustdar et al. (2016b) emphasise on the need of a cloud/edge based public utility service ecosystem where consumers, producers and brokers can manage privacy protected distributed micro-transactions (e.g. blockchain). Babar and Arif (2017) propose a smart city architecture for big data analytics with three generic layers - data acquisition & aggregation, data computation and processing and application specific decision making. This layered architecture is common with others like Khan et al. (2015); Rathore et al. (2016a); Zhang et al. (2016). Survey indicates that data pre-processing is a common and essential step to prepare raw data into more appropriate format for analysis and decision making. For data preparation and analysis, different techniques or approaches are used e.g. pre-processing or data normalisation (min-max approach); data aggregation for analytics e.g. Hadoop model, Kalman Filter for speeding up the processing by separating data noise, Round Robin and Least Slack Time for scheduling and load balancing. Hadoop or map-reduce based model is

useful for static data but with increasing real-time data and need to process and to make near real-time decisions emphasise using Spark and Storm. Such platforms can support governance of individual themes such as energy (Carli et al., 2015) or multi-disciplinary applications (Khan et al., 2015; Rathore et al., 2016a).

For good city governance, monitoring and real-time decision making can play an important role as highlighted by K. R. Malik et al. (2018) where raw data is transformed in JSON and RDF for semantic inferencing and querying. For instance, Eirinaki et al. (2018) use a multi-tier architecture (Angular JS, NodeJS, Apache Mahout, MySQL with interfaces with Socrata Open Data Platform) in Amazon EC2 cloud to handle online building permit applications. This not only facilitates citizens to easily communicate with complex planning permission processes but it also performs cross-departmental data collection and data analytics to enable city planners get the likelihood of building constructions in specific parts of a city. The monitoring is also important for policy making in cities where policy provenance through data capturing and processing can be used for policy analytics (Javed et al., 2018).

Citizen centred data analytics can provide necessary intelligence to city authority in provisioning of user related public services. Zhang et al. (2016) introduced a multi-layered framework for citizen centred urban governance and decision making. Their proposed multi-layered platform consists of three layers: data merging i.e. combining citizen-centred data with other urban sources; knowledge discovery i.e. citizen profile building through citizen participation, statistical and machine learning methods; and, decision making i.e., using ontology, data mining, Bayesian Net techniques. However citizen centred data challenges such as managing heterogenous structured, unstructured or semi-structured citizen data requires various techniques such as: cleaning, splitting, translating, merging, sorting, and validating data. Aligning user profiles with urban data and analysing it using data mining and machine learning techniques, the platform is able to deliver user-centred public services resulting in effective urban governance. Likewise Aguilera et al. (2017) also build citizen centric apps by exploiting open data in IES Cities platform. In IES platform, citizens are considered as prosumers that is consume and produce data resulting in better provisioning of public services to citizens and decision making. The platform uses heterogenous urban data in formats CSV, RDF, JSON, relational and provides toolset to enable developers to

develop new urban apps. Motta et al. (2014) present City Feed —- a crowdsourcing system that covers transactional and analytical modules to measure the maturity of crowd-based city governance. The platform uses an intelligence engine to perform collective analysis on the crowdsource data based on well-defined dimensions. The above literature shows that it is becoming increasingly important in open governance to engage with citizens and take their opinion in urban planning and decision making e.g. smarticipate platform where citizens can vote and comment on an existing proposal; create alternative proposals and can get processed rule-based feedback (Khan, Dambruch, et al., 2017).

Though existing literature emphasise on the use of urban Big data analytics, alignment between technology innovation, human dimension i.e. socio-techno perspective and city management processes; evaluation in real city environment highlight lack of integrated and/or cross-departmental data analytics for smart governance. For example, Chatfield and Reddick (2017) investigated 311 on demand services (i.e. non-emergency police and government services) of a large US city and impact of big data analytics. Their investigation reveals use of big data analytics by different departments but its usage on a wider city scale was not clearly evident. As a result authors proposed process-based approach as an institutional mechanism for embedding big data analytics use in critical governance processes to co-create public values and creating a culture of analytics driven governance in public authorities. This also can establish smart and connected communities in contributing to data driven governance. In this respect Sun et al. (2016) argue that smart and connected communities (SCC) need remembering past i.e. preservation, living in present i.e livability and need to plan for future i.e. sustainability. Therefore, the ubiquitous IoTs, smart sensors, cloud computing and big data analytics can play an important role in promoting SCC. Author, presents a layered platform 'TreSight' for data sensing, context building, event handling & processing, big data analytics for gaining insights, secure identity management (OAuth) and providing context-aware services e.g. Tourism & cultural heritage service mobile app.

Pan et al. (2016a) explore impact of big data in smart city intelligence for infrastructure support, urban governance, public services and economic and industrial development in China. Authors present an overview of urban big data with multi-dimensional hierarchy, integrity and correlation which require systems to adopt targeted methods for data lifecycle

i.e. data acquisition (e.g. filtering), pre-processing (cleaning, filling, normalisation, quality check etc), processing & analysing (i.e. linear analysis, non-linear analysis, factor analysis, sequential analysis, variable curve analysis, bivariate statistics, linear regression, etc.), categorisation and inter-relation analysis (i.e. support vector machine, Naive Bayes, random forest, logistic regression, etc.), uncovering patterns, rules and knowledge (i.e. AI neural network, genetic algorithms, cross-media algorithm), and value generation (interactive and visual information) with automated routines. This suggests that data driven smart city analytics and governance have common steps where numerous techniques or approaches can be applied as part of the system architecture.

Above literature review clearly indicates that data analytics is becoming common in smart city governance. Most of the frameworks or platforms have common functionalities such as data collection, preparation, processing, analysis and application specific decision-making. Citizen-centric data and integrated (i.e. cross-departmental) data analytics for urban governance is challenging and requires data preparation, coordination & alignment between technology and city processes, and political willingness is needed in order to gain operational feasibility for data driven smart city governance. Most of the literature covers different technologies e.g. IoTs, Sensors, Big Data, Cloud, Edge, Fog, Semantics, Data management, etc; however not all of these literature articles covered security and privacy issues of data driven urban governance.

Environment

An important pillar of smart cities is the smart environment (Chatterjee & Kar, 2015; Vinod Kumar, 2015), an umbrella term for a number of different areas. As mentioned previously, one of the challenges that smart cities face is to reduce its carbon footprint to mitigate climate change (van Dijk & Teuben, 2015). This can be done in a number of different ways, e.g. by switching to a renewable source of energy such as solar, by reducing air pollution and thus, improving air quality and even by motivating people to switch to alternative modes of transport, resulting in fewer numbers of cars on the road. Due to this reason, the theme of smart environment can cover a number of different topics. We focus specifically on tools and techniques that involve the use of big data analytics. After searching

relevant literature and analysing their abstracts and years of publication, we short listed 10 papers for detailed review.

The recent emergence of IoT has resulted in a deluge of data in many fields, including smart cities. Air quality monitoring in particular is an area that has benefited greatly from IoT (Molka-Danielsen et al., 2017; Tahat et al., 2018; Wang & Eick, 2017). The presence of a large number of sensors that acquire readings every few minutes generates a significant amount of data that needs to be stored, processed and analysed efficiently and in near real-time. This is why big data analytics has proven to be a popular choice for processing and analysing such data. For example, GSMA (2018) has recently produced a report citing the advantages of using big data analytics coupled with the IoT to monitor air quality data as well as the opportunities available.

Hu et al. (2018) propose ClimateSpark, an in-memory data analytics framework for processing climate-related big data. ClimateSpark is based on Apache Spark (Ganelin et al., 2016) and includes ClimateRDD, a data model for representing climate data that allows for more efficient processing of the data that is based on the Resilient Distributed Dataset (RDD) format. ClimateRDD chunks up the spatiotemporal climate data both temporally and spatially into logical units. This allows for better parallel I/O as well as provides advanced data querying capabilities. As ClimateSpark uses Spark for its core data processing capabilities, all processing is performed in-memory, which also greatly speeds up the processing time.

Another framework that targets geo-spatial data mining in order to detect changes in the ozone concentration in various geographic locations was presented by Wang and Eick (2017). The framework includes pre-processing of geo-spatial data, a density-based data mining algorithm and a visualisation algorithm. The framework also introduces a change pattern discovery technique to detect dynamic change patterns in geo-spatial data, particularly to detect changes in the ozone concentration. A post-processing analysis technique is also presented that automatically identifies interesting spatial clusters output by the data mining algorithm.

Pei et al. (2018) propose the ATWDP framework that is used for analysing solar cell welding data. It offers a number of services such as early warning, real-time monitoring and knowledge discovery. The framework has three layers; the enterprise layer, the sensor layer

and the batch layer. Data is collected from solar cells using sensors embedded in them in the sensor layer. The enterprise layer is for interfacing with the users and is also used for data acquisition. The batch layer processes the collected data using MapReduce combined with an SVM-DS (Cortes & Vapnik, 1995).

Tahat et al. (2018) present a cloud-based big data analytics framework to analyse environmental data. Data is collected through sensors and sent to a central cloud platform for storage. A MongoDB database stores the user authentication information while an InfluxDB is used to store the time series monitoring data. The Orange data mining framework is used to analyse the data. The framework also has an associated iOS mobile app to access and visualise the data. Molka-Danielsen et al. (2017) present two case studies to investigate how big data analytics can be used along with wireless sensor networks to monitor air quality in workplace environments. The data is collected from wireless sensors and stored in a MicroSD card for offline processing. Even though their aim was to study how big data analytics can be used to monitor air quality the nature of their data is such that it cannot be considered big data. Lu et al. (2011) propose a framework to perform distributed big data analytics on climate data using RapidMiner (Arimond, Kofler, & Shafait, 2010). The authors also implement a data conversion tool to convert the NetCDF (Rew & Davis, 1990) climate data into RapidMiner native format.

Zheng et al. (2013) present U-Air, a co-learning framework that uses two separate classifiers to predict the air quality of a particular geographic location. The framework uses a spatial classifier that takes into account spatially-related features such as road networks, distance to an existing station, points of interests etc. The other classifier is a temporal classifier that takes into account temporally-related features such as humidity, temperature and traffic flow. Le (2018) adopt a slightly different approach. They use a Convolutional Neural Network to process "image-like" air pollution concentration maps and classify the pollution according to hazard level in real time. On the other hand they use a combination of a Neural Network and Long Short-term Memory network to take into account historical time series air pollution data and current local factors such as temperature, humidity, wind direction etc to predict future air pollution. Zhu, Cai, Yang, and Zhou (2018) present yet another method for prediction air pollution concentrations. They use a machine learning

approach that uses regularisation to predict hourly air pollution levels based on historical meteorological data as well as current air pollution data.

A major concern in contemporary analytics-based systems is the privacy and security of user data. With the implementation of the GDPR, this has become of even greater importance. Anisetti et al. (2018) propose a semi-automated privacy-aware framework solution to ensure compliance with the GDPR. The authors propose a policy definition process that produces a machine-readable policy using deontic logic. Modelling policies this way allows them to generate machine-executable workflows that in turn produce evidence using big data analytics for decision-making. This deontic model also incorporate privacy requirements which allows the system to automatically check if the defined policy conforms to GDPR requirements.

Energy

Energy plays a very important role in improving the living environment of Smart Cities by making environment cleaner, recycling of energy, lower energy consumption, lower emissions to reduce pollution and by having lower energy bills (TRANSFORM Project Consortium, 2015). With the recent advancements in IoT and Cloud technologies, energy systems are being made to work more intelligently based on the data collected from their operations. According to Report Linker, smart energy systems will grow at a CAGR of 14.91% in terms of revenue by 2020 which will contain artificial intelligent devices to work without human interference (Clare, 2016). This part of the paper focusses on the data analytics related to Energy to achieve the above mentioned objectives.

An architecture for energy data analytics is proposed by M. Malik et al. (2017) for making data easily reproducible and accessible for broader purposes like usability, transformation and with stream support. The proposed architecture has 3 main components for data collection, data storage and data distribution. For data collection a schema consisting of Type and Units of value measured, Geo-coordinates of the location of measurement with frequency of measurement and time stamp is proposed to store data from different sensor types in a single table. For data distribution, a publisher-subscriber based architecture is proposed to support both batch and stream data processing. For data storage, it recommends to use

Polyglot persistence architecture based on multiple database technologies according to the requirements for application or components of application.

The imbalance between increasing energy demands and dwindling energy supply has intensified over period of time. A Big data and data mining technology based solution is presented by Horban (2016) to increase the energy efficiency. Power failures are common in developing countries but, in recent years these have become more frequent in developed countries as well. Ji, Wei, and Poor (2017) address the issue of failure in Energy systems and resilience using data analytics especially caused by natural disaster like weather. The data about failures, recoveries and weather variables are collected and analysed to find the resilience level of infrastructure and services. The system also provides modelling technique to provide a pertinent role of guiding data analytics like what measurements to use and quantities to estimate etc. The system tries to answer the questions related to Outage management and failure prediction from the location aware data collected automatically with high accuracy. For the support of study, data analytics were performed on a data set generated at a national level across the whole USA to quantify the resilience of the energy infrastructure and services under severe weather conditions.

The energy sharing through a substantial enhancement of its operational efficiency rather than relying on establishing new energy networks has become of utmost important to enable governments to meet their expenses in energy sector. The paper in F. Li et al. (2018) discusses sharing the energy systems in two major categories according to Customer Flexibility and Network Flexibility. It is encouraged to share the energy with flexibility on Peer-to-Peer (P2P) energy trading in local energy markets supported by Shared Network Access (SNA). SNA aims to entertain flexible demands in a cost effective manner. The paper presents this idea based on Big data analytics as a key enabler to inform and enable users and market participants to share energy based on established prices and a match on flexibility in demand.

From the literature review on energy, big data and analytics can be seen as big enablers to gain insights which can help in solving the issues of energy systems e.g. achieving lower consumption and lower emissions in order to make environment more cleaner and lowering the energy bills through sharing or behavioural change.

Economy

Healthy economy is a critical factor for cities to function properly and provide services to citizens. Economic forecasting aims to provide future prediction/forecast of the economy on the basis of certain values. These values can be high and low prices of goods, services (inflation), employment etc. Precise measurements about the economic requirements of populations importantly influence researchers and government. These measurements manipulate governments to allocate the limited resources and present the foundation about the need to global world for taking decision and improving decisions about the livelihoods of human. The economic quality and quantity of data is much improved in many developing countries but still insufficient in some countries. This needs effort to identify and involvement in the poor areas. Economic Forecast is mainly used by businesses and governments to assist them in deciding their budgets, multi-year plans and strategy for the upcoming year. Stock exchange market specialists use forecasts to help them estimate the valuation of a company and its stock. In the age of Big Data, it can be used to forecast economic aspects, by producing meaningful knowledge for future. Based on the current data and information, machine learning algorithms can be applied to achieve better forecast results. Athey (2017) discusses the impact of machine learning and its approaches on Economic policies and also comparison of traditional approaches with machine learning. Future trends and contribution of Machine Learning with economics are also discussed in this paper. The author has also discussed state of the art research in econometric and statistic, and then performs contrast with traditional approaches. Some of the problems related to decision making are explained using different computational approaches. Following are the examples discussed by Athey (2017):

Hip Replacement Surgery: Prediction model is very helpful in making decisions regarding the health of a person. It is difficult for poor people to arrange money for their treatment. If one can guess what will happen in the future, then decisions can be made accordingly. Hip replacement example is used to predict whether to go for a surgery of an older person or not based on the health conditions.

Health Code Violation: In restaurants, health code violations are used by the health inspector in order to evaluate the performance of the restaurant and to rate them

according to their food quality, health and safety. Prediction Models are used in this paper to predict health code violations in restaurants.

Nyman, Ormerod, Smith, and Tuckett (2014) describe a new approach to economic forecasting, which is based on the availability of "Big Data". The main aim is to develop an algorithm to analyse text data which is given in different forms for example news feed, email, office legal documents recording to extract time series of sentiment analysis that are similar in nature, that might forecast aspects for the economy. Deep Learning and Machine learning methods are widely used to predict poverty in many countries including countries from Africa (Blumenstock, 2016; Gebru et al., 2017; Jean et al., 2016). These approaches produce better results for predicting and identifying poverty and wealth. Satellite images contain a lot of information. But the images are unstructured. The biggest challenge is to extract meaningful information from it. Transfer learning is being used (Jean et al., 2016) to overcome these challenges. This method has some noise but it is able to obtain proxy and train a deep learning model for poverty. The transfer learning approach consists of three steps. Firstly, a convolutional neural network (CNN) (LeCun, Bengio, & Hinton, 2005) has been pretrained on a large dataset of ImageNet. ImageNet consists of labelled images. This model is able to learn features such as edges and corners. After retrieving the images from CNN model, the CNN again fine tunes the model in a new task. CNN model is used as a feature extractor with survey data to predict the poverty estimations. Most important market driver in this case study is poverty prediction using socio-economic data in African regions especially Uganda, Nigeria, Tanzania, Malawi. As it is difficult to scale-up conventional data-collection approaches, alternative approach to data-collection is using novel sources of "passively collected data" such as social media, mobile phone networks and satellites. Economic characteristics of populations influences a lot of research and accurate measurement of such data can shape critical decisions for the governments. Livelihood can be improved by allocating resources in optimal fashion and by tracking progress.

The regression technique can also be used to predict the future income of a person from his/her current financial statistics and the expected growth rate in the financial value of that person. There are different methodologies to compute the income of a person from his/her current financial statistics. The most common way of future value (of person) prediction is

26

modelling and analysis of Earnings Dynamics relevant to the Consumption Dynamics and the environment uncertainty and change for a person. There are a lot of heterogeneity in these models but these models are producing good results. These dynamics can be analysed by different regression techniques.

Druedah and Munk-Nielsen (2018) provided a linked regression tree based algorithm to analyse income and liabilities dynamics. The analysis is done on a data of Danish males from the age of 30 to 60. The attributes for the analysis are income, assets and debts. The income is the yearly earning of that person and the assets are housing equity, financial assets including the bank account balance, stocks, bonds and funds, and the debts that he has to pay to any authority including bank loans. The Linked Regression Tree (LRT) algorithm is used to estimate the mean and the variance of the levels of log income over a life cycle of 30 years from the ages of 30 to 60. LRT algorithm converts the data into income groups and the problem of regression into the problem of classification. The income of a person at any age (in future) is dependent upon his current income and the growth rate in the income. The Markov model (a probability based model) is used to predict the income group of a person at any age with current income group as a current state and the growth in income as a transition. One of the extra finding other then the actual prediction that was found from the results is that, a person uses his 10% income on welfare activities. They used different measures like mean square error of group specific income-based on the age and 5-fold cross validation.

Machine learning incorporates empirical resolution of tasks instead of constituting a hard-wired rule which would map input x to a predicted value y. This essential quality can be applied in the field of econometrics where parameter estimation is key in the estimation of relationship between y and x. However, machine learning is more synchronous with finding predicted values instead of parametric values. Mullainathan and Spiess (2017) discuss how machine learning can be tuned to become an integral part in the econometrics and what pitfalls to avoid in overcoming the inherent problem of overfitting. This research has used 10,000 records of owner-occupied unit's data from 2011 metropolitan sample of American Housing Survey. It has compared multiple classifiers such as regression trees and LASSO with simple procedural estimation functions such as least squares to find why machine learning

is a more portent tool in econometrics. Although, random forests perform much better in prediction on out-of-sample data than ordinary least squares it also has a tendency to overfit. This resonates with most machine learning approaches wherein high dimensionality can both be a bane and a boon: to have a large set of features does offer good generalization but it really is overfitting on a sample data which then translates to poor performance on out-of-sample. This is where regularization becomes essential – to apply methods to tune the algorithm in such a way as to increase its prediction power on unseen data. This empirical tuning includes procedures such as cross-validation and using these sub-samples to apply regularization parameters. Finally, the parameter with best average performance is selected (Angrist & Pischke, 2008). This is being used in LASSO which is most used by economists. This incorporates a quadratic loss function, class of linear functions and a regularizer as sum of coefficients' absolute value (Bekker, 1994). Conclusively, the efficiency of machine learning lies in not deducing the rules but inducing them instead. Thus, instead of rigidly applying functions uniformly on all types of data, the data is being allowed to make informative decision. Nevertheless, machine learning works in conjunction with theory driven approaches and helps in making more informed decisions in choosing which variables/functions to select in order to best make predictions on unseen data.

Smart cities and sharing economy are the current topics of prototype development and research, where different issues are discussed such as trend of current urbanization, economically facing situation which people are facing for many years. Gori, Parcu, and Stasi (2015) have discussed about sharing economy and smart cities, their common features, similarities and differences. Authors have suggested some policies to Government for welfare of masses. Smart cities as defined by European Commission are a system where people use power, energy, resources to improve their living standards and smart communication means and information use help to achieve this goal. A smart city emerges when government participates for improving living standards by investing in people, transport, ICT, communication infrastructure, sustainable economic growth while managing natural resources. Both smart city and sharing economy depends upon improved utilization of resources that can be either money, assets, time or services. Data protection and security are also important terms. Camerer (2017) has expressed the idea of connecting Machine Learning with behavioural eco-

nomics. They perform the experiments on the basis of evaluating bargaining and risk choice. The main theme is how artificial intelligence (AI) can be used to assemble preferences of predictions regarding the unfamiliar product.

In essence, AI and machine learning can easily transform the economic structure and thus, manipulation can be minimized. Moreover ML and AI will focus policy problems and new research areas will born and process will be changed due to automation. ML will automate the routine task of data analysis and so on the economist will be able to enjoy the credibility of empirical work.

Transport

As the world is changing, cities are growing and urban population is rising, the need for transport for goods and people is increasing. Hence we need more mobility but in smarter way. Being a significant component of modern economy, transportation constitute 6 to 12% of Gross Domestic Product (GDP) in many developed countries. It also has great impact on people's life as estimated around 10 to 15% of household expenditure is of transportation and around 8% time of daily life is spent on travel by people. Though transportation has greatly improved people's lives; few costly problems remain unsolved including congestion, traffic accidents, air pollution and climate change due to fuel consumption and particulate emission of the vehicles. Continuous efforts have been done in the past to solve these mentioned problems but little significant success is made. Recently, Transportation has become a hot topic in Internet of Things (IoT) area and is considered part of the solution to mentioned problems (Gottbehüt, 2016).

Based on IoT technology, companies across various travel and transportation industry sectors - including railways, aerospace industry, airlines, freight logistics and others - have started capturing the data from almost all kind of systems or events in the public or private clouds. With advancement of cloud technology to store data and sophisticated data analysis techniques to extract insights from data at high speed and accuracy, organizations are making better knowledgeable decisions which was not achievable before (Wedgwood & Howard, 2015). This way data itself has achieved strategic and competitive asset importance and data analytics tools and technologies are contributing to solve the problems of congestion,

accidents and climate change.

This part of the survey focuses to show the advancement in data analytics technology, issues and available solutions. From the literature search, following papers were selected to be presented in this survey, mainly, based on their relevance to the topic.

As data is of key importance to perform analytic, Shukla and Champaneria (2017) present different ways to collect data in order to provide better transportation management. For counting the number of devices entering or exiting the road, different technologies are presented including pneumatic tubes, inductive loops and vehicle's detection via video. Pneumatic tubes are installed on the surface of the road to count the vehicles by the compressed burst of air by the air switch when a vehicle passes near to the tube. Inductive Loops count the passing vehicles by detecting the vehicle from the change in magnetic field. Besides these all, a modern video based technology is used now a days. The vehicles can be identified and counted by successfully identifying them in video by the intelligent software which can identify the different objects and separate vehicles from various background objects. The paper also talks about collecting GPS based data from mobile phones or tracker fitted in the car to collect information about vehicle speed or traffic density. Different types of sensors are also presented in the paper to measure CO2 using Nondispersive Infrared Sensors (NIDR), visibility using Fog sensors, presence of vehicle in some specific lane using Ultrasonic sensors and Magnetometer sensor for vehicle classification, detection and speed estimation.

To reduce the congestion and road accidents, a big data analytic based model is presented by Abberley, Gould, Crockett, and Cheng (2017) to capture the semantics of road accidents and congestion using ontology. The paper uses Big data quantitative information to provide qualitative information about traffic ontology using Machine learning techniques. The four steps model tries to formulate a conceptual model of congestion in step 1 to develop an ontology for congestion considering the impact of road accidents. In step 2, congestion causing dimensions due to accidents are identified such as journey time and traffic volume. In step 3, the Big data sources relevant to identified dimensions are identified to collect the data for analysis. At the end, in the step 4, the identified dimensions and data from identified data sources is analysed to identify patterns which can lead to better decision makings to reduce congestion. Machine learning algorithm Clustering is used to analyse different traffic

and congestion situations in different week days for different times of day to show the impact of schools and office travels.

Another important objective to meet from Smart Transportation is to have better Climate. Electric Vehicles are seen as an important component to achieve this objective. B. Li, Kisacikoglu, Liu, Singh, and Erol-Kantarci (2017) present a survey on data analytic techniques based on Big data for reducing emissions for greener smart city is presented with focus on smart grids and electric vehicles. Different Hadoop based and web based data analytic techniques are presented. Battery consumption and charging meters are managed by Hadoop R statistical package and stream processing using Hadoop Pig scripts respectively. For web based solutions, decision tree algorithms , J48 and M5 algorithms are used on Weka platform to analyse the data collected from NY Independent System Operators (NYISO).

Literature has also discussed the systems to manage the overall traffic systems. For example Sharif et al. (2017) present a low cost Smart Traffic System (STS) is also proposed to deploy traffic updates instantly while Rizwan, Suresh, and Babu (2016) propose a traffic management system that tries to solve the traffic problems specifically in India.

Rizwan et al. (2016) propose a Smart Traffic Management System (STMS) to solve the problem of traffic management in India especially congestion. The low cost STMS aims to provide better service by traffic indicators which are updated instantly depending upon the traffic conditions. It suggests to deploy sensor in the middle of the road at every 500 meters on highways to collect different types of data from the vehicles. The big data collected is analysed for traffic density and predict the future condition for traffic considering the number of vehicles entering the road and capacity of the road. The traffic conditions information and possible alternative routes are sent to the drivers on a mobile phone application developed for the proposed system. The drivers can take alternative routes to avoid the congested areas.

The overall literature review shows that data analytic techniques based on machine learning, Big data and artificial intelligence in general are trying to solve the long present issues of traffic management successfully. With this progress, it looks like that congestion, road accidents and traffic management problems will be solved in future smart cities using IoT and data analytic technologies to a greater extent.

DISCUSSION AND CONCLUSIONS

Table 5 summarises challenges and possible solutions presented in the literature regarding smart city big data analytics. The main challenges faced for each surveyed theme are:

Governance

For smart governance, city administrators, planners, policy and decision makers can benefit from evidence-based collective information e.g. performance indicators, public value, financial gains, citizen centric public services, etc., based on which they can plan effective policies and perform impact assessment. However, most of the applications are built for vertical thematic silos due to lack of inter-departmental collaboration, data standardisation and data sharing. This suggests that there is a need for Big data strategy for smart city where privacy-preserved open data, crowd-sourced or citizen-centred data, data from public different agencies or/and departments can be securely gathered, harmonised or integrated and processed for collective information intelligence to support timely decision making. The use of disruptive technologies including IoT, sensors, 2D/3D GIS, Cloud, Edge, Fog, Hyper ledgers, Human-Computer Interaction e.g. interactive web, AR/VR, distributed computing models such as service orientation, map-reduce, real-time stream processing, context-awareness and processing models using existing and new algorithms in spatial analysis, machine learning, AI, data mining are key enablers for smart governance. However, smart governance solutions can only be beneficial if there is willingness in city administrations to transform their processes and adopt these solutions. Therefore, there is need for close collaboration between various stakeholders e.g. domain experts, business organisations, citizens and IT experts to design, develop and adopt smart governance services.

Environment

Smart environments are enabled by the use of IoT-based sensor networks. For example, pollution, wind, temperature, humidity etc sensors can be deployed across a city to get a real-time overview of the weather situation as well as predict weather changes. Such monitoring systems typically collect data on an hourly basis and analyse it to produce an

overview. The use of IoT in such a way produces a large amount of data that needs to be processed in near real-time on an almost hourly basis. Machine learning is by far the most common technique for performing these kinds of analyses. There are also a number of other challenges that arise from these sorts of monitoring systems, e.g. how to represent the data such that analyses can be performed efficiently on it. Environment monitoring systems also typically suffer from bad quality data and big data analysis techniques can again help to preprocess such data for further processing. Future challenges in smart environments include exploring the use of deep learning-based techniques to make the aforementioned monitoring systems more efficient and effective.

Energy

Smart Energy sector has been also the major focus of the research in IoT and Smart Cities research. It is going to improve the energy supply for household supplies according to requirements and two-way transportation of energy will save lot of power as well. Smart meters has been also a vital component of research in Smart energy domain. They are going to support the efficient usage of energy and saving money to the people by calculating the two-way transportation of energy. It will also make the life of meter readers easy by sending the billing information directly to the billing server. However, more research should be focusing on standardization of data models and analytics to enable users and businesses to provide energy on established prices in way that is more flexible. Also the industries producing the products for Smart Grids and other relevant sector, need to emphasize more on the generation of standards for data collection, storage and transmission of data.

Economy

Correct estimates of inflation, spending, employment etc are some of the most important aims of economic forecasting (Nyman et al., 2014). Economic recession of 2008 / 09 was not predicted by the researchers despite data availability (Nyman & Ormerod, 2016). One of the factors is the incorrect approximation of the household income process (Druedah & Munk-Nielsen, 2018). Good estimation of the household income has a long history in

economics.

Another challenge is how artificial intelligence can interact with behavioural economics. Behavioural economics is defined as the analysis of the willpower, natural limits on computation, and self-interest, and the effects of those boundaries for economic analysis (Camerer, 2017). It is vital to study how AI technology used in business can both exploit and overcome human limits.

Big Data does provide increasing opportunities to learn knowledge that might be useful to the above task. Current, economic trends produce more volume, variety and velocity of data more than before. The main challenge is to extract useful information from big data in order to get meaningful data. In recent years, a lot of machine learning techniques have been proposed to accomplish that task. Supervised and unsupervised machine learning techniques can be quite useful here since these methods can be used to create variables that can be useful in standard econometric analyses. Curse-of-Dimensionality is another challenge in analysing economic data. Curse-of-Dimensionality arises when number of samples are less than the number of variables. Dimensionality reduction and Feature selection techniques are quite useful to solve curse-of-dimensionality problem. For emerging economies, big data analytics can also be useful to support global manufacturing and supply chain innovation by improving human decision-making and data transparency (Tan, Ji, Lim, & Tseng, 2017).

Transportation

Transportation has been the major focus of researchers for smart cities studies. Autonomous vehicles has revolutionize the industry and are going to be available in next 2–7 years according to different sources. The data analytics are helping aerospace, airlines, trains and cars industries to solve their different long existing problems. As discussed in the paper in Transport section, many projects are done to solve the problems of congestion, car park, congestion and pollution. It is analysed from literature review that most of the products in transport area are commercial and all products are following their own structures and approach to store and analyse the data. There is need to develop the standards for data processing specially collection and storage structures for data. Therefore, there is a need that standardization bodies, researchers and industry should work together to standardize these

things to accelerate the development of more products from more vendors in this domain.

Current and Future Trends

Gartner (n.d.) provides a methodology to assess current and future technology trends. The key stages for a technology identified by this methodology are:

Peak of Inflated Expectations: early stages where the hype of a technology is at its peak and there's lots of publicity surrounding it.

Trough of Disillusionment: experiments and implementations begin to fail and interest wanes in the technology.

Slope of Enlightenment: understanding of how a technology can benefit enterprises and systems improves and concrete applications begin to appear.

Plateau of Productivity: a technology is adopted in the mainstream and becomes broadly applicable in the market.

Technologies important for enabling smart cities such as deep learning, machine learning and blockchains are currently at the peak of their inflated expectations (Panetta, 2018). In the coming years the AI technologies are expected to enter their trough of disillusionment. This will entail reduced interest in these technologies followed by deeper understanding of their potential and finally robust and wider adoption. Blockchain is still a relatively new phenomenon so it is not expected to mature until the next 5 to 10 years. Technologies like edge computing, IoT etc are still in the initial stages of their hype cycle and so are not expected to reach maturity for another 10 years or so. This indicates that these are the technologies of the future with the greatest potential for investment, research and innovation. These trends are also borne out by our survey, where research involving IoT, edge computing, augmented reality etc is still only just emerging. In addition to these new technologies such as Virtual Assistants, Digital Twin, Edge and Fog AI etc are new trends that have emerged in the past year that are important for smart city applications. AI, Data mining, machine learning are the core methods which enable all these technologies that have the potential to reshape the smart city landscape and open up new avenues for research and innovation in the coming years.

Table 5: Overview of big data analytics challenges and solutions in smart cities.

Themes	Smart City Issues		
	Features	**Challenges**	**Solutions**
Governance	Data Models	• Big data strategy (Kourtit & Nijkamp, 2018; PWC, n.d.) • open data (Barns, 2017) • crowd-sourced or citizen centred data (Gil-Garcia et al., 2016; Khan, Dambruch, et al., 2017; Motta et al., 2014; Peters-Anders et al., 2017) • data from public agencies, in-situ and remote sensing data (Aguilera et al., 2017; Lv et al., 2018; Pan et al., 2016b; Zhang et al., 2016) • data harmonisation or compatibility (K. R. Malik et al., 2018) • horizontal and vertical integration and processing (Aguilera et al., 2017; Babar & Arif, 2017; Carli et al., 2015; Eirinaki et al., 2018; PWC, n.d.)	• metadata, data standardisation and lifecycle management (Aguilera et al., 2017) • semantics for interoperability
	Computing Models	• technology development (e.g. cloud, fog/edge IoT, sensors, GIS, VR) (Eirinaki et al., 2018; Lv et al., 2018; Silva et al., 2018; Sun et al., 2016) and its acceptance across different city departments (Chatfield & Reddick, 2017; Gil-Garcia et al., 2016; PWC, n.d.) • data fusion, exploration, statistical and predictive analysis (Babar & Arif, 2017; Pan et al., 2016b; Zhang et al., 2016) • distributed computing models for batch and real-time data processing (Babar & Arif, 2017; Lv et al., 2018) • added value (Dustdar et al., 2016b), increased productivity (Barns, 2017; Motta et al., 2014), contextual (Rathore et al., 2016a) or evidence-based decision making (Khan et al., 2015), policy making (Kourtit & Nijkamp, 2018) and provenance (Javed et al., 2018)	• technology rollout strategy (PWC, n.d.) • machine learning, data mining, AI, spatial analysis (Lv et al., 2018; Zhang et al., 2016) • city data platforms or city dashboards (Barns, 2017; Carli et al., 2015; Kourtit & Nijkamp, 2018) • cloud (Lv et al., 2018), edge, IoT (Dustdar et al., 2016b) Hadoop, spark, storm (Babar & Arif, 2017), ontology (Zhang et al., 2016) • measurable performance indicators, public value (Chatfield & Reddick, 2017), impact assessment (Pan et al., 2016b) e.g., financial gains (Meijer & Bolívar, 2016), efficiency (PWC, n.d.), citizen centric public services (Zhang et al., 2016) etc.

Themes	Smart City Issues		
	Features	Challenges	Solutions
	Security and Privacy	• Individual privacy and confidentiality (Sun et al., 2016) • Directives and legislations e.g. EU's General Data Protection Regulations (Anisetti et al., 2018)	• security techniques e.g. authentication, authorisation, anonymisation, encryption, etc. • compliance to legislations
	Market Drivers	• Commercial products • open source or research-oriented solutions	• Commercial vendors e.g. IBM, Microsoft, Fujitsu, HP, etc.; Research and academia e.g. open source initiatives such as commonly used Apache tools for distributed analytics (Aguilera et al., 2017; Babar & Arif, 2017; Rathore et al., 2016a)
Environment	Data Models	• Aggregating and storing data from sensors • Data my be missing or inaccurate • Data is often of a spatial and temporal nature, increasing its complexity	• Develop models to manage the complex spatiotemporal nature of the data (Hu et al., 2018) • Store data in distributed and time series databases • Develop models that optimise parallel I/O for faster data access
	Computing Models	• Data is voluminous and needs to be handled processed in near real-time • The spatiotemporal nature of data needs to be taken into account during processing	• New, multi-faceted computing models needs to be developed to process spatiotemporal data (Ghaemi et al., 2018; Le, 2018, etc.) • Distributed data processing mechanisms need to be used.
	Security and Privacy	• Data processing and handling needs to be compliant with privacy-related regulations such as GDPR.	• Introduce methodology, tools and techniques to ensure compliance (Anisetti et al., 2018).
	Market Drivers	• Commercial products • Open source frameworks and tools	

Themes	Smart City Issues		
	Features	Challenges	Solutions
Economy	Data Models	• Analyse Income and Liabilities Dynamics (Druedah & Munk-Nielsen, 2018) • Human expectations (Camerer, 2017) • Health care challenges • Machine Learning for Econometric (Mullainathan & Spiess, 2017) • Economic Recession data (Nyman & Ormerod, 2016)	• Data Collection from Danish males of 30 to 60 (Druedah & Munk-Nielsen, 2018) • Collecting behaviour data (Camerer, 2017) • Elderly people surgery information (Athey, 2017) • Collection of owner occupied unit (Mullainathan & Spiess, 2017), early 90s, early 20s and $2007 - 2009$ great recession data
	Computing Models	• Accurately predict future income of a person (Druedah & Munk-Nielsen, 2018) • Patient life expectancy, Economic Forecasting, Predicting Economic Recession, Predict Poverty	• Linked Regression Tree • Machine Learning • Gradient Descent, Lasso Regression, Directed Algorithm Text Analysis (Mullainathan & Spiess, 2017) • Ordinary Least Squares Regression, Deep Learning, Transfer Learning (LeCun et al., 2005)
	Security and Privacy	• Confidentiality, Safety	• security techniques e.g. authentication, authorisation, anonymisation, encryption, etc. • compliance to legislations
	Market Drivers	• Commercial Intelligent products • open source or research-oriented solutions	• Insurance companies • Banking Sector • Academics
Energy	Data Models	• Data Transfer and Management, • Data Reproducibility (M. Malik et al., 2017) • Efficient data comparison from different sources. • Big Data Analytics and Data Mining (Horban, 2016) • Achieving Resilience by Data Analytics (Ji et al., 2017) • Big Data Analytics for Flexible Energy Sharing (F. Li et al., 2018)	• Defined Schema building (M. Malik et al., 2017) • Data Processing Framework in different layers (Horban, 2016)

Themes	Smart City Issues		
	Features	**Challenges**	**Solutions**
Transport	Computing Models	• Usability, Reproducibility, Transformation and Stream support (M. Malik et al., 2017) • Complexity, Velocity and Diversity of Big Data Streams to compare different data streams (Horban, 2016)	• Distributed data processing based on publish/subscribe model (M. Malik et al., 2017) • Machine Learning and Data Mining (Horban, 2016)
	Safety and Privacy	• Authentication, Integrity and Safety	• Identity Verification, • Hashes or Message Authentication Codes (MACs) and • Monitoring for Physical Injuries
	Market Drivers	• Commercial Intelligent Products • Open Source Frameworks and Tools	
	Data Models	• Analysing data from different sources • Counting the passing vehicles (Shukla & Champaneria, 2017) • Big Data Analytics to reduce the Congestion (Abberley et al., 2017) • Big Data Analytics for Greener affect (B. Li et al., 2017) • Traffic Management and Congestion (Rizwan et al., 2016)	• Semantics from various sources and for various types of data (Shukla & Champaneria, 2017) • Ontology Definitions • Machine Learning (Shukla & Champaneria, 2017)
	Computing Models	• Interpretation of Data collected on various angles and from different sources (Shukla & Champaneria, 2017) • Qualitative information generation from quantitative data interpretation	• Sensors based infrastructure to analyse different types of data and process (Shukla & Champaneria, 2017) • Clustering and Machine learning techniques (Abberley et al., 2017)
	Safety and Privacy	• Privacy, Confidentiality, Authentication and Safety	• Anonymity, • Encryption/Decryption, • Identity Verification and • Monitoring and Protection for Physical Injuries
	Market Drivers	• Commercial Intelligent Products • Open Source Frameworks and Tools	

SUMMARY

In this paper we have discussed the challenges faced by smart cities and the key role data mining, machine learning and statistical methods can play to enable intelligent solutions for different applications. We present a structured data mining-based literature review methodology which we use to study the role of big data analytics smart cities and present major challenges and applied data mining and machine learning methods. Some of the major points discussed are:

1) Smart cities are a fertile area where big data analytics has a significant role to play.

2) There are several data mining, machine learning and statistical methods commonly applied in smart city big data analytics. These are: Support Vector Machines, Decision Tree, Nearest Neighbour, Bayesian inference, Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM), Autoencoders (AEs), Generative Adversarial Networks (GANs), and Deep Belief Network (DBNs), SVM-DS, LRT.

3) The advent of IoT has further generated the need for big data analytics techniques as an unprecedented level of data is being generated and collected.

4) Big data is applicable to many different areas of smart cities such as governance, environment, economy, transportation, etc.

5) Big data analytics can support governance through enabling the mining of citizen preferences and opinions collected from a number of sources such as social media, polls, online comments etc. In addition, cross-thematic and cross-departmental data analysis is essential for smart governance. Smart governance needs to consider both socio-technical aspects and need proper mechanisms to introduce data driven strategies in city processes.

6) Big data analytics can also support evidence-based decision-making by integrating and bringing together multi-sectoral data and helping to extract information out of it.

7) In the field of smart environments big data analytics can help to collect, manage and analyse big data collected from a number of different sources such as environmental sensors, historical climate and mobility data, predict future environmental scenarios

and visualise existing situations etc.

8) In smart transportation big data analytics can help to solve the problems of traffic congestion by finding alternate routes earlier, reducing pollution by electric vehicles, traffic parking and in time provision of medical services after accidents. Location based data and vehicles data should be analysed by advanced analytical techniques to solve the traffic problems.

9) In smart economy big data can help to manage and analyse massive amounts of historical economic data and analyse and predict future economic scenarios. In this regard economic forecasting is a major application.

10) For smart energy big data analytics can support the living of people in Smart Cities by providing cleaner environment, recycling of energy, lower energy consumption and emissions to reduce pollution and lowering the energy bills for homes and businesses. The consumers will also be generating energy and will contribute to better energy management.

11) From a technological standpoint a number of big data technologies have proven to be very useful for support smart cities. For example, distributed storage solutions such as Amazon EC2 and HDFS help to manage the massive amounts of smart city data. Distributed data processing platforms such as Hadoop and Microsoft Azure help to process the smart city data. NoSQL databases like MongoDB and time series databases like InfluxDB lend themselves to storing the collected and processed smart city data.

From what we have been able to discern, so far newer technologies like deep learning, digital twin, virtual assistants, Edge AI are still in their early stages and offer significant opportunities for future research and development. For example, deep learning could allow smart city representatives to analyse the already collected data on a more deeper level and to improve the efficiency of the solutions that have already been developed. Virtual assistants can apply AI, data mining and machine learning methods to make cross-thematic city data more usable and enable city stakeholders to query public services in an efficient way. Also, less resource intensive algorithms on edge nodes would be highly desirable in smart city applications. Security, privacy and trust are significant issues going forward and must be

taken into account by any new data analytic solution. For example, data assimilation must not violate privacy of individuals and transparency of algorithms can add trust on the new information.

FUNDING INFORMATION

REFERENCES

Abberley, L., Gould, N., Crockett, K., & Cheng, J. (2017, Sept). Modelling road congestion using ontologies for big data analytics in smart cities. In *2017 international smart cities conference (isc2)* (p. 1-6). doi: 10.1109/ISC2.2017.8090795

Abu-Elkheir, M., Hassanein, H. S., & Oteafy, S. M. A. (2016, Sept). Enhancing emergency response systems through leveraging crowdsensing and heterogeneous data. In *2016 international wireless communications and mobile computing conference (iwcmc)* (p. 188-193). doi: 10.1109/IWCMC.2016 .7577055

Aguilera, U., Peña, O., Belmonte, O., & López-de Ipiña, D. (2017). Citizen-centric data services for smarter cities. *Future Generation Computer Systems*, *76*, 234–247.

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

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton, New Jersey, USA: Princeton university press.

Anisetti, M., Ardagna, C., Bellandi, V., Cremonini, M., Frati, F., & Damiani, E. (2018, May). Privacy-aware Big Data Analytics as a service for public health policies in smart cities. *Sustainable Cities and Society*, *39*, 68–77.

Antonić, A., Marjanović, M., Pripužić, K., & Žarko, I. P. (2016). A mobile crowd sensing ecosystem enabled by cupus: Cloud-based publish/subscribe middleware for the internet of things. *Future Generation Computer Systems*, *56*, 607–622.

Arimond, A., Kofler, C., & Shafait, F. (2010). Distributed pattern recognition in RapidMiner. In *Proceedings of rapidminer community meeting and conference.*

Athey, S. (2017). *Impact of machine leaning on economics.* The national bureau of economic research.

Babar, M., & Arif, F. (2017). Smart urban planning using big data analytics to contend with the interoperability in internet of things. *Future Generation Computer Systems*, *77*, 65–76.

Barkham, R., Bokhari, S., & Saiz, A. (2018). *Urban big data: City management and real estate markets* (Tech. Rep.). GovLab.

Barns, S. (2017). Smart cities and urban data platforms: Designing interfaces for smart governance. *City, Culture and Society*.

Bassoo, V., Ramnarain-Seetohul, V., Hurbungs, V., Fowdur, T. P., & Beeharry, Y. (2018). Big data analytics for smart cities. In N. Dey, A. E. Hassanien, C. Bhatt, A. S. Ashour, & S. C. Satapathy (Eds.), *Internet of things and big data analytics toward next-generation intelligence* (pp. 359–379). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-60435-0_15 doi: 10.1007/978-3-319-60435-0_15

Bawany, N. Z., & Shamsi, J. A. (2015). Smart city architecture: Vision and challenges. *International Journal of Advanced Computer Science and Applications*, *6*(11), 246–255.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, *62*(3), 657–681.

Bibri, S. E. (2017). The iot for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society*.

Bibri, S. E., & Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustainable Cities and Society*, *31*, 183–212.

Bilal, K., Khalid, O., Erbad, A., & Khan, S. U. (2018). Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers. *Computer Networks*, *130*, 94–120.

Blumenstock, J. E. (2016). Fighting poverty with data. *Science*, *353*(6301), 753–754. Retrieved from http://science.sciencemag.org/content/353/6301/753 doi: 10.1126/science.aah5217

Bolívar, M. P. R. (2015). Smart cities: Big cities, complex governance? In *Transforming city governments for successful smart cities* (pp. 1–7). Springer.

Camerer, C. F. (2017). *Artificial intelligence and behavioral economics.* In Economics of Artificial Intelligence.

Carli, R., Albino, V., Dotoli, M., Mummolo, G., & Savino, M. (2015). A dashboard and decision support tool for the energy governance of smart cities. In *Environmental, energy and structural monitoring systems (eesms), 2015 ieee workshop on* (pp. 23–28).

Chatfield, A. T., & Reddick, C. G. (2017). Customer agility and responsiveness through big data analytics for public value creation: A case study of houston 311 on-demand services. *Government Information Quarterly*.

Chatterjee, S., & Kar, A. K. (2015, Aug). Smart cities in developing economies: A literature review and policy insights. In *2015 international conference on advances in computing, communications and informatics (icacci)* (p. 2335-2340). doi: 10.1109/ICACCI.2015.7275967

Chauhan, S., Agarwal, N., & Kar, A. K. (2016, Jun). Addressing big data challenges in smart cities: a systematic literature review. *info*, *18*(4), 73–90. Retrieved from http://dx.doi.org/10.1108/info-03-2016-0012 doi: 10.1108/info-03-2016-0012

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: literature review and challenges. *International Journal of Distributed Sensor Networks*, *11*(8), 431047.

Clare. (2016, July 27). *Global smart energy market 2016-2020* [Online]. Retrieved 5th May 2018 from https://www.prnewswire.com/news-releases/global-smart-energy-market-2016-2020-300304980.html.

Cortes, C., & Vapnik, V. (1995, Sep 01). Support-vector networks. *Machine Learning*, *20*(3), 273–297. Retrieved from https://doi.org/10.1023/A:1022627411411 doi: 10.1023/A:1022627411411

Digiesi, S., Facchini, F., Mossa, G., Mummolo, G., & Verriello, R. (2015). A cyber - based dss for a low carbon integrated waste management system in a smart city. *IFAC-PapersOnLine*, *48*(3), 2356 - 2361. Retrieved from http://www.sciencedirect.com/science/article/pii/S2405896315006795 (15th IFAC Symposium onInformation Control Problems inManufacturing) doi: https://doi.org/10.1016/j.ifacol.2015.06.440

Docherty, I., Marsden, G., & Anable, J. (2017). The governance of smart mobility. *Transportation Research Part A: Policy and Practice*.

Druedah, J., & Munk-Nielsen, A. (2018). Higher-order income dynamics with linked regression trees. In *2nd conference on dynamic structural models: Methodology and applications of structural dynamic models and machine learning.* Department of Economics, University of Copenhagen.

Dustdar, S., Nastic, S., & Scekic, O. (2016a). A novel vision of cyber-human smart city. In *Hot topics in web systems and technologies (hotweb), 2016 fourth ieee workshop on* (pp. 42–47).

Dustdar, S., Nastic, S., & Scekic, O. (2016b, Oct). A novel vision of cyber-human smart city. In *2016 fourth ieee workshop on hot topics in web systems and technologies (hotweb)* (p. 42-47). doi: 10.1109/HotWeb.2016.16

Edmondson, V., Cerny, M., Lim, M., Gledson, B., Lockley, S., & Woodward, J. (2018). A smart sewer asset information model to enable an 'Internet of Things' for operational wastewater management. *Automation in Construction*, *91*, 193 - 205. Retrieved from http://www.sciencedirect.com/science/article/pii/S0926580517306945 doi: https://doi.org/10.1016/j.autcon.2018.03.003

Eirinaki, M., Dhar, S., Mathur, S., Kaley, A., Patel, A., Joshi, A., & Shah, D. (2018). A building permit system for smart cities: A cloud-based framework. *Computers, Environment and Urban Systems*.

El Mendili, S., El Idrissi, Y. E. B., & Hmina, N. (2016). Benchmarking study on smart city data analytics. In *Information science and technology (cist), 2016 4th ieee international colloquium on* (pp. 841–846).

enerquire. (2017, September 15). *Disruptive innovation in the energy sector – definition & application to renewables* [Blog post]. Retrieved from https://www.enerquire.com/blog/disruptive-innovations

`-in-the-energy-sector-definition-application`.

Faieq, S., Saidi, R., Elghazi, H., & Rahmani, M. D. (2017). C2iot: A framework for cloud-based context-aware internet of things services for smart cities. *Procedia Computer Science*, *110*, 151–158.

Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., & Mankodiya, K. (2018). Towards fog-driven iot ehealth: Promises and challenges of iot in medicine and healthcare. *Future Generation Computer Systems*, *78*, 659 - 676. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0167739X17307677` doi: https://doi.org/10.1016/j.future.2017.04.036

Ganelin, I., Orhian, E., Sasaki, K., & York, B. (2016). *Spark: Big data cluster computing in production.* Wiley. Retrieved from `https://books.google.co.uk/books?id=SDa7CwAAQBAJ`

Gartner. (n.d.). *Gartner hype cycle.* Retrieved 22nd August, 2018, from `https://www.gartner.com/en/research/methodologies/gartner-hype-cycle`

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, *114*(50), 13108–13113. Retrieved from `http://www.pnas.org/content/114/50/13108` doi: 10.1073/pnas.1700035114

Ghaemi, Z., Alimohammadi, A., & Farnaghi, M. (2018, Apr 20). LaSVM-based big data learning system for dynamic prediction of air pollution in Tehran. *Environmental Monitoring and Assessment*, *190*(5), 300. Retrieved from `https://doi.org/10.1007/s10661-018-6659-6` doi: 10.1007/s10661-018-6659-6

Gil-Garcia, J. R., Zhang, J., & Puron-Cid, G. (2016). Conceptualizing smartness in government: An integrative and multi-dimensional view. *Government Information Quarterly*, *33*(3), 524–534.

Gori, P., Parcu, P. L., & Stasi, M. L. (2015). *Smart cities and sharing economy.* Robert Schuman Centre for Advanced Studies, European University Institute, Fiesole, Italy, Paper No. RSCAS 2015/96.

Gottbehüt, C. (2016, February 1). *Connecting to the smart future, smart transportation* (Report). Huawei. Retrieved 6th May 2018 from `http://www.ey.com/Publication/vwLUAssets/Connecting-the-cities-of-the-future/$FILE/Connecting-the-cities-of-the-future.pdf`.

GSMA. (2018, February). *Air quality monitoring using IoT and big data* (Report). Retrieved 2nd May 2018 from `https://www.gsma.com/iot/wp-content/uploads/2018/02/iot_clean_air_02_18.pdf`.

H., F. J., & Vipin, K. (2014). A big data guide to understanding climate change: The case for theory-guided data science. *Big Data*, *2*(3), 155-163. Retrieved from `https://doi.org/10.1089/big.2014.0026` (PMID: 25276499) doi: 10.1089/big.2014.0026

Habibzadeh, H., Boggio-Dandry, A., Qin, Z., Soyata, T., Kantarci, B., & Mouftah, H. T. (2018). Soft sensing in smart cities: Handling 3vs using recommender systems, machine intelligence, and data analytics. *IEEE Communications Magazine*, *56*(2), 78–86.

Haneem, F., Kama, N., Ali, R., & Selamat, A. (2017, 09). Applying data analytics approach in systematic literature review: Master data management case study. In *The 16th international conference on*

intelligent software methodologies, tools, and techniques (somet_17). Kitakyushu, JAPAN. doi: 10.3233/978-1-61499-800-6-705

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748–758.

Horban, V. (2016, Aug). A multifaceted approach to smart energy city concept through using big data analytics. In *2016 ieee first international conference on data stream mining processing (dsmp)* (p. 392-396). doi: 10.1109/DSMP.2016.7583585

Hu, F., Yang, C., Schnase, J. L., Duffy, D. Q., Xu, M., Bowen, M. K., ... Song, W. (2018, June). ClimateSpark: An in-memory distributed computing framework for big climate data analytics. *Computers & Geosciences*, *115*, 154–166.

Ismail, K. A., Majid, M. A., Zain, J. M., & Bakar, N. A. A. (2016, Oct). Big data prediction framework for weather temperature based on mapreduce algorithm. In *2016 ieee conference on open systems (icos)* (p. 13-17). doi: 10.1109/ICOS.2016.7881981

Iyengar, S., Lee, S., Irwin, D., Shenoy, P., & Weil, B. (2018). Watthome: A data-driven approach for energy efficiency analytics at city-scale. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 396–405). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/3219819.3219825 doi: 10.1145/3219819.3219825

Javed, B., Khan, Z., & McClatchey, R. (2018). An adaptable system to support provenance management for the public policy-making process in smart cities. *Informatics*, *5*(1). doi: 10.3390/informatics5010003

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. Retrieved from http://science.sciencemag.org/content/353/6301/790 doi: 10.1126/science.aaf7894

Ji, C., Wei, Y., & Poor, H. V. (2017, July). Resilience of energy infrastructure and services: Modeling, data analytics, and metrics. *Proceedings of the IEEE*, *105*(7), 1354-1366. doi: 10.1109/JPROC.2017.2698262

John Walker, S. (2014, Jan). Big data: A revolution that will transform how we live, work, and think. *International Journal of Advertising*, *33*(1), 181–183. Retrieved from http://dx.doi.org/10.2501/IJA-33-1-181-183 doi: 10.2501/ija-33-1-181-183

Kamilaris, A., Gao, F., Prenafeta-Boldu, F. X., & Ali, M. I. (2016, Dec). Agri-iot: A semantic framework for internet of things-enabled smart farming applications. In *2016 ieee 3rd world forum on internet of things (wf-iot)* (p. 442-447). doi: 10.1109/WF-IoT.2016.7845467

Khan, Z., Anjum, A., Soomro, K., & Tahir, M. A. (2015). Towards cloud based big data analytics for smart future cities. *Journal of Cloud Computing*, *4*(1), 2.

Khan, Z., Dambruch, J., Peters-Anders, J., Sackl, A., Strasser, A., Fr^hlich, P., ... Soomro, K. (2017). Developing knowledge-based citizen participation platform to support smart city decision making: The smarticipate case study. *Information*, *8*(2). doi: 10.3390/info8020047

Khan, Z., Pervez, Z., & Abbasi, A. G. (2017). Towards a secure service provisioning framework in a smart city environment. *Future Generation Computer Systems*, *77*, 112 - 135. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X17313687   doi: https://doi.org/10.1016/j.future.2017.06.031

Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities–an experiment on stockholm. *Land Use Policy*, *71*, 24–35.

Kousiouris, G., Akbar, A., Sancho, J., Ta-shma, P., Psychas, A., Kyriazis, D., & Varvarigou, T. (2018). An integrated information lifecycle management framework for exploiting social network data to identify dynamic large crowd concentration events in smart cities applications. *Future Generation Computer Systems*, *78*, 516 - 530. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X17300055 doi: https://doi.org/10.1016/j.future.2017.07.026

Le, D. (2018, April). Real-time air pollution prediction model based on spatiotemporal big data. *arXiv.org*.

LeCun, Y., Bengio, Y., & Hinton, G. (2005). Deep learning. *Nature*, *521*.

Li, B., Kisacikoglu, M. C., Liu, C., Singh, N., & Erol-Kantarci, M. (2017, NOVEMBER). Big data analytics for electric vehicle integration in green smart cities. *IEEE Communications Magazine*, *55*(11), 19-25. doi: 10.1109/MCOM.2017.1700133

Li, F., Li, R., Zhang, Z., Dale, M., Tolley, D., & Ahokangas, P. (2018, May). Big data analytics for flexible energy sharing: Accelerating a low-carbon future. *IEEE Power and Energy Magazine*, *16*(3), 35-42. doi: 10.1109/MPE.2018.2796358

Linder, L., Vionnet, D., Bacher, J.-P., & Hennebert, J. (2017). Big building data - a big data platform for smart buildings. *Energy Procedia*, *122*, 589 - 594. Retrieved from http://www.sciencedirect.com/science/article/pii/S1876610217329582 doi: https://doi.org/10.1016/j.egypro.2017.07.354

Liu, Z., Yan, S., Cao, J., Jin, T., Tang, J., Yang, J., & Wang, Q. (2018, Dec). A bayesian approach to residential property valuation based on built environment and house characteristics. In *2018 ieee international conference on big data (big data)* (p. 1455-1464). doi: 10.1109/BigData.2018.8622422

Lu, S., Li, R. M., Tjhi, W. C., Lee, K. K., Wang, L., Li, X., & Ma, D. (2011). A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data. In *Cloud computing technology and science (cloudcom), 2011 ieee third international conference on* (pp. 618–622).

Lv, Z., Li, X., Wang, W., Zhang, B., Hu, J., & Feng, S. (2018). Government affairs service platform for smart city. *Future Generation Computer Systems*, *81*, 443 - 451. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X17311391   doi: https://doi.org/10.1016/j.future.2017.08.047

Malik, K. R., Sam, Y., Hussain, M., & Abuarqoub, A. (2018). A methodology for real-time data sustainability in smart city: Towards inferencing and analytics for big-data. *Sustainable Cities and Society*, *39*, 548 - 556. Retrieved from http://www.sciencedirect.com/science/article/pii/S2210670717305735 doi: https://doi.org/10.1016/j.scs.2017.11.031

Malik, M., Tabone, M., Chassin, D. P., Kara, E. C., Guha, R. V., & Kiliccote, S. (2017, Oct). A common data architecture for energy data analytics. In *2017 IEEE international conference on smart grid communications (smartgridcomm)* (p. 417-422). doi: 10.1109/SmartGridComm.2017.8340736

Meijer, A., & Bolívar, M. P. R. (2016). Governing the smart city: a review of the literature on smart urban governance. *International Review of Administrative Sciences*, *82*(2), 392-408. Retrieved from https://doi.org/10.1177/0020852314564308 doi: 10.1177/0020852314564308

MGI. (2013). *Disruptive technologies: Advances that will transform life, business, and the global economy* (Tech. Rep.). McKinsey Global Institute. (Available online [Last accessed: 25th April 2018])

Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018, Fourthquarter). Deep learning for iot big data and streaming analytics: A survey. *IEEE Communications Surveys Tutorials*, *20*(4), 2923-2960. doi: 10.1109/COMST.2018.2844341

Mohammadi, M., & Al-Fuqaha, A. I. (2018). Enabling cognitive smart cities using big data and machine learning: Approaches and challenges. *CoRR*, *abs/1810.04107*. Retrieved from http://arxiv.org/abs/1810.04107

Mohammadi, M., Al-Fuqaha, A. I., & Oh, J. (2018). Path planning in support of smart mobility applications using generative adversarial networks. *CoRR*, *abs/1804.08396*. Retrieved from http://arxiv.org/abs/1804.08396

Mohanty, S. (2015, June 10). *The four essential V's for a big data analytics platform.* Retrieved 28th April 2018, from http://dataconomy.com/2015/06/the-four-essentials-vs-for-a-big-data-analytics-platform/

Molka-Danielsen, J., Engelseth, P., Olesnanikova, V., Sarafin, P., 2017, & Zalman, R. (2017). Big Data Analytics for Air Quality Monitoring at a Logistics Shipping Base via Autonomous Wireless Sensor Network Technologies. In *2017 5th international conference on enterprise systems (es)* (pp. 38–45). IEEE.

Motta, G., You, L., Sacco, D., & Ma, T. (2014). City feed: A crowdsourcing system for city governance. In *Service oriented system engineering (sose), 2014 ieee 8th international symposium on* (pp. 439–445).

Mullainathan, S., & Spiess, J. (2017, May). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87-106. Retrieved from http://www.aeaweb.org/articles?id=10.1257/jep.31.2.87 doi: 10.1257/jep.31.2.87

Nguyen, K., Le, T., Nguyen, T. D., Phung, D., & Webb, G. I. (2018). Robust bayesian kernel machine via stein variational gradient descent for big data. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*. Retrieved from http://dx.doi.org/10.1145/3219819.3220015 doi: 10.1145/3219819.3220015

Niu, L., Tang, Q., Veeraraghavan, A., & Sabharwal, A. (2018, June). Learning from noisy web data with category-level supervision. In *2018 ieee/cvf conference on computer vision and pattern recognition* (p. 7689-7698). doi: 10.1109/CVPR.2018.00802

Nyman, R., & Ormerod, P. (2016). *Predicting economic recessions using machine learning algorithms.* Cornell University Library (Working Paper).

Nyman, R., Ormerod, P., Smith, R., & Tuckett, D. (2014, April). Big data and economic forecasting: A top-down approach using directed algorithmic text analysis. In *Ecb workshop on big data for forecasting and statistics.* Frankfurt.

Osman, A. M. S., Elragal, A., & Bergvall-Kåreborn, B. (2017). Big data analytics and smart cities: A loose or tight couple? In *International conference on connected smart cities 2017 (csc 2017), lisbon, 20-22 july 2017.*

Pan, Y., Tian, Y., Liu, X., Gu, D., & Hua, G. (2016a). Urban big data and the development of city intelligence. *Engineering*, *2*(2), 171–178.

Pan, Y., Tian, Y., Liu, X., Gu, D., & Hua, G. (2016b). Urban big data and the development of city intelligence. *Engineering*, *2*(2), 171 - 178. Retrieved from http://www.sciencedirect.com/science/article/pii/S2095809916309456 doi: https://doi.org/10.1016/J.ENG.2016.02.003

Panetta, K. (2018, August). *Top 5 trends emerge in the Gartner hype cycle for emerging technologies, 2018.* Retrieved 9th April 2019, from https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/

Pei, F.-Q., Li, D.-B., & Tong, Y.-F. (2018, May). Double-layered big data analytics architecture for solar cells series welding machine. *Computers in Industry*, *97*, 17–23.

Pérez-González, D., & Díaz-Díaz, R. (2015, feb). Public services provided with ICT in the smart city environment: The case of Spanish cities. *Journal of Universal Computer Science*, *21*(2), 248–267. (http://www.jucs.org/jucs_21_2/public_services_provided_with)

Peters-Anders, J., Khan, Z., Loibl, W., Augustin, H., & Breinbauer, A. (2017). Dynamic, interactive and visual analysis of population distribution and mobility dynamics in an urban environment using the mobility explorer framework. *Information*, *8*(2). doi: 10.3390/info8020056

Pettit, C., Bakelmun, A., Lieske, S. N., Glackin, S., Hargroves, K. C., Thomson, G., ... Newman, P. (2018). Planning support systems for smart cities. *City, Culture and Society*, *12*, 13 - 24. Retrieved from http://www.sciencedirect.com/science/article/pii/S1877916617302175 (Innovation and identity in next generation smart cities) doi: https://doi.org/10.1016/j.ccs.2017.10.002

Preda, S., Oprea, S.-V., Bâra, A., & Belciu (Velicanu), A. (2018). Pv forecasting using support vector machine learning in a big data analytics context. *Symmetry*, *10*(12). Retrieved from http://www.mdpi.com/2073-8994/10/12/748 doi: 10.3390/sym10120748

PWC. (n.d.). *Data driven cities* (Tech. Rep.). Retrieved 26th April 2018 from https://www.pwc.ru/en/government-and-public-sector/assets/ddc_eng.pdf.

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016a). Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks*, *101*, 63 - 80. Retrieved from http://www.sciencedirect.com/science/article/pii/S1389128616000086 (In-

dustrial Technologies and Applications for the Internet of Things) doi: https://doi.org/10.1016/
j.comnet.2015.12.023

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016b). Urban planning and building smart cities based
on the internet of things using big data analytics. *Computer Networks*, *101*, 63–80.

Rathore, M. M., Paul, A., Hong, W.-H., Seo, H., Awan, I., & Saeed, S. (2017). Exploiting iot and big data
analytics: Defining smart digital city using real-time urban data. *Sustainable Cities and Society*.

Rew, R., & Davis, G. (1990). Netcdf: an interface for scientific data access. *IEEE computer graphics and
applications*, *10*(4), 76–82.

Rizwan, P., Suresh, K., & Babu, M. R. (2016, Oct). Real-time smart traffic management system for
smart cities by using internet of things and big data. In *2016 international conference on emerging
technological trends (icett)* (p. 1-7). doi: 10.1109/ICETT.2016.7873660

Rouse, M. (2017, March). *Big data analytics.* Retrieved 28th April 2018, from https://
searchbusinessanalytics.techtarget.com/definition/big-data-analytics

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for
value-creation. *Information Processing & Management*.

Sharif, A., Li, J., Khalil, M., Kumar, R., Sharif, M. I., & Sharif, A. (2017, Dec). Internet of things
— smart traffic management system for smart cities using big data analytics. In *2017 14th interna-
tional computer conference on wavelet active media technology and information processing (iccwamtip)*
(p. 281-284). doi: 10.1109/ICCWAMTIP.2017.8301496

Shukla, S. N., & Champaneria, T. A. (2017, Feb). Survey of various data collection ways for smart trans-
portation domain of smart city. In *2017 international conference on i-smac (iot in social, mobile,
analytics and cloud) (i-smac)* (p. 681-685). doi: 10.1109/I-SMAC.2017.8058265

Silva, B. N., Khan, M., & Han, K. (2017). Integration of big data analytics embedded smart city architec-
ture with restful web of things for efficient service provision and energy management. *Future Gener-
ation Computer Systems*. Retrieved from http://www.sciencedirect.com/science/article/pii/
S0167739X17305174 doi: https://doi.org/10.1016/j.future.2017.06.024

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

Simmhan, Y. (2018). Big data and fog computing. In S. Sakr & A. Zomaya (Eds.), *Encyclopedia of big
data technologies* (pp. 1–10). Cham: Springer International Publishing. Retrieved from https://
doi.org/10.1007/978-3-319-63962-8_41-1 doi: 10.1007/978-3-319-63962-8_41-1

Singh, D., & Mohan, C. K. (2018, Dec). Projection-svm: Distributed kernel support vector machine for
big data using subspace partitioning. In *2018 ieee international conference on big data (big data)*
(p. 74-83). doi: 10.1109/BigData.2018.8622478

Stephenson, D. (2013, January 18). *7 big data techniques that create business value.* Retrieved 28th
April 2018, from https://www.firmex.com/thedealroom/7-big-data-techniques-that-create

`-business-value/`

Sun, Y., Song, H., Jara, A. J., & Bie, R. (2016). Internet of things and big data analytics for smart and connected communities. *IEEE Access*, *4*, 766–773.

Susmitha, K., & Jayaprada, S. (2017, Aug). Smart cities using big data analytics. *International Research Journal of Engineering and Technology (IRJET)*, *4*(8), 1615–1620.

Tahat, A., Aburub, R., Al-Zyoude, A., & Talhi, C. (2018). A smart city environmental monitoring network and analysis relying on big data techniques. *Proceedings of the 2018 International Conference on Software Engineering and Information Management - ICSIM2018*. Retrieved from `http://dx.doi.org/10.1145/3178461.3178464` doi: 10.1145/3178461.3178464

Tan, K. H., Ji, G., Lim, C. P., & Tseng, M.-L. (2017). Using big data to make better decisions in the digital economy. *International Journal of Production Research*, *55*(17), 4998-5000. doi: 10.1080/00207543.2017.1331051

Terroso-Saenz, F., Gonzalez-Vidal, A., Cuenca-Jara, J., & Skarmeta, A. F. (2017, Oct). An open architecture for iot data analytics in smart cities. In *2017 global wireless summit (gws)* (p. 235-239). doi: 10.1109/GWS.2017.8300468

TRANSFORM Project Consortium. (2015, March 3). *Work package1: 'becoming a smart energy city, state of the art and ambition'* (Tech. Rep.). Retrieved from `http://urbantransform.eu/wp-content/uploads/sites/2/2015/07/WP1-revised-final-report.march-2015.pdf`.

Treadwell, J. (2017, July 11). *Disruptive trends in transport – smart cities, citizens and intelligent connections.* Presentation at the 8th World Bank – Korea Transport Institute Joint Workshop, Washington DC, USA. (Available online: `http://pubdocs.worldbank.org/en/986981501515858559/WB-Jane-Treadwell-Disruptive-Trends-in-Transport.pdf`)

van Dijk, A., & Teuben, H. (2015). *Smart cities: How rapid advances in technology are reshaping our economy and society* (Tech. Rep.). Deloitte. (Available online: `https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/public-sector/deloitte-nl-ps-smart-cities-report.pdf` [Last access: 25th April 2018])

Vinod Kumar, T. M. (2015). E-governance for smart cities. In T. M. Vinod Kumar (Ed.), *E-governance for smart cities* (pp. 1–43). Singapore: Springer Singapore. Retrieved from `https://doi.org/10.1007/978-981-287-287-6_1` doi: 10.1007/978-981-287-287-6_1

Wang, S., & Eick, C. F. (2017, September). A data mining framework for environmental and geo-spatial data analysis. *International Journal of Data Science and Analytics*, *5*(2-3), 83–98.

Wedgwood, K., & Howard, R. (2015, November). *Big data and analytics in travel and transportation* (Report). IBM. Retrieved 6th May 2018 from `http://www-07.ibm.com/au/pdf/Travel_Transportation.pdf`.

Yin, C., Xiong, Z., Chen, H., Wang, J., Cooper, D., & David, B. (2015). A literature survey on smart cities. *Science China Information Sciences*, *58*(10), 1–18.

Zaman, H. B., Ahmad, A., Hamid, N. A., Sin, A. K., Hussain, A., Hannan, M. A., & Saad, H. M. (2017, Oct). Integrated smart neighborhood framework and application to sustain an innovative digital economy in the 4ir and big data era. In *2017 3rd international conference on science in information technology (icsitech)* (p. 7-12). doi: 10.1109/ICSITech.2017.8257077

Zaouali, K., Rekik, R., & Bouallegue, R. (2018, June). Deep learning forecasting based on auto-lstm model for home solar power systems. In *2018 ieee 20th international conference on high performance computing and communications; ieee 16th international conference on smart city; ieee 4th international conference on data science and systems (hpcc/smartcity/dss)* (p. 235-242). doi: 10.1109/HPCC/SmartCity/DSS.2018.00062

Zhang, Q., Zhang, X., Zhang, Q., Shi, W., & Zhong, H. (2016). Firework: Big data sharing and processing in collaborative edge environment. In *Hot topics in web systems and technologies (hotweb), 2016 fourth ieee workshop on* (pp. 20–25).

Zheng, Y., Liu, F., & Hsieh, H.-P. (2013). U-air. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*. Retrieved from http://dx.doi.org/10.1145/2487575.2488188 doi: 10.1145/2487575.2488188

Zheng, Y., Yi, X., Li, M., Li, R., Shan, Z., Chang, E., & Li, T. (2015). Forecasting fine-grained air quality based on big data. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. Retrieved from http://dx.doi.org/10.1145/2783258.2788573 doi: 10.1145/2783258.2788573

Zhu, D., Cai, C., Yang, T., & Zhou, X. (2018). A machine learning approach for air quality prediction: Model regularization and optimization. *Big Data and Cognitive Computing*, *2*(1). Retrieved from http://www.mdpi.com/2504-2289/2/1/5 doi: 10.3390/bdcc2010005

Zotano, M. A. G., & Bersini, H. (2017). A data-driven approach to assess the potential of smart cities: the case of open data for brussels capital region. *Energy Procedia*, *111*, 750–758.