A Survey on Smart Cities, Big Data, Analytics, and Smart Decision-making

Towards an analytical framework for decision-making in smart cities

Marius Rohde Johannessen, Lasse Berntzen,
Department of Business, History and Social Sciences
University of South-Eastern Norway
Kongsberg, Norway
e-mail: {lasse.berntzen; marius.johannessen}@usn.no

Rania El-Gazzar
Department of Business and Law
University of South-Eastern Norway
Kongsberg, Norway
e-mail: rania.el-gazzar@usn.no

***Abstract*—Smart decision making is based on data combined with analytics to improve decision-making. This paper examines several application areas of smart cities, and related data sources used for decision-making. Further, we present a review of analytical techniques used in earlier studies. In many cases, systems may make decisions on their own. Such autonomous systems may play an essential role in the development of smart cities. In other cases, the data can be combined with historical data or other open data sources to play a role as the foundation for decision-making. Our findings are presented as an analytical framework, which will be used for further empirical studies into this domain.**

Keywords-smart decision-making; smart cities; big data; sensors; analytics; autonomous systems.

I. Introduction

This article is an expanded version of an earlier conference paper presented at ICDS 2018 [1], and offers an analytical framework for smart or (intelligent) decision-making in the context of smart cities. The framework is based on a review of literature, white papers and news sources covering the topic, as well as empirical data from a study on air quality monitoring. The analytical framework shows areas in need of further study and forms the basis for future research projects.

The analytical framework shows areas in need of further study and as such forms a research agenda for (big) data analysis in a smart city context. The target audience for this work is mainly Information Systems (IS) researchers and practitioners.

Smart decision-making uses a systematic approach to data collection and applies logical decision-making techniques instead of using intuition, generalizing from experience, or trial and error.

“Smart cities” is a multifaceted concept and has been defined in many different ways; more than 100 definitions of smart cities have been analyzed by the International Telecommunication Union (ITU)’s focus group on smart sustainable cities [2][3]. The mandatory requirement for smart cities is to improve quality of life and achieve sustainable development (economic, social, and environmental) through the use of Information and Communications Technology (ICT) and intelligence [4]. Definitions emphasized the technological aspect of a smart city as being "a technologically interconnected city" or Internet of Things (IoT) using big data is promoted to achieve the efficiency and intelligence in managing cities' resources [5][6].

A smart city is a city that is characterized as “instrumented, interconnected, and intelligent” [7][8][9]. This can be conceptualized as three layers, as shown in Figure 1.

These characteristics are enabled by the use of ICT, which constitute the heart of a smart city [10]. The “instrumentation” layer does data acquisition through sensor-based systems that provide real-time data through sensors, meters, and cameras, but also from social media and open data sources. The instrumentation layer enables capturing and filtering data from various sources for timely response. The inputs from the instrumentation layer are integrated and transformed into event-related information at the “interconnection” layer to provide rich insights for decision-making. The interconnection layer provides all forms of collaboration among people, processes, and systems to enable a holistic view supporting decision-making. At the “intelligence” layer, business intelligence and analytics are applied to the information provided by the interconnection layer and other city-relevant data and, then, the analyzed information is visualized to understand the city requirements and city policies, hence, make informed decisions and take actions. The intelligence layer is focused on deep discovery, analyses, and forecasting. These three layers that build up the smartness in a smart city are constructed by smart technology solutions and ICT infrastructure, such as IoT, big data, and the Internet.

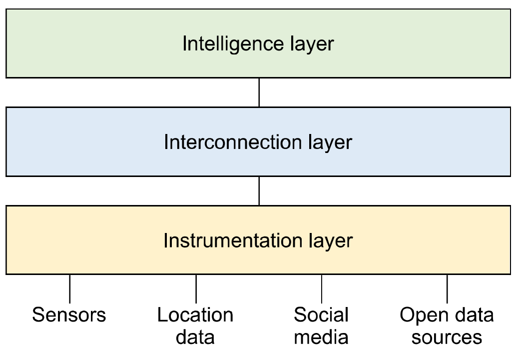

Figure 1. Three-layer model.

Regarding the intelligence layer that is concerned with decision-making, a review of studies on smart city and decision-making resulted in nine articles. This indicates that smart city and decision-making is an area that deserves further investigation on how to make a significant impact from big data [11].

In this article, we elaborate on smart or intelligent decision-making in the context of smart cities. Smart decision-making relies on data and analytics to make better decisions. By using autonomous systems, the decisions can be implemented in real time. Human intervention can be reduced to oversee the decisions and take over if the system is malfunctioning.

The primary focus of this article is on the instrumentation and intelligence layers, and the data sources and analytical techniques used for decision-making. The data is refined through the interconnection layer and processed by the intelligence layer to enable decision-making. The three-level model provides a systematic approach to collecting facts and applying logical decision-making techniques, instead of generalizing from experience, intuition (guessing), or trial and error.

The rest of the article is structured as follows: Section II discusses methodology. Section III focuses on the instrumentation layer, including identification of common data sources. Section IV describes some selected smart city application areas. Section V presents an overview of relevant analytical techniques. Section VI presents our analytical framework. Section VII contains our conclusion, some limitations, and ideas for future work.

II. Methodology

The purpose of this article is to begin exploring how common application areas of smart cities use, analyze and visualize data. Data analysis and visualization are essential for decision-making and intelligence in smart cities [7]-[9]. However, our literature review reveals little research in this area.

Figure 2 shows how data is analyzed and visualized. The analytics typically stores data for future use, e.g., for predictions. The visualization is used for human decision-making.

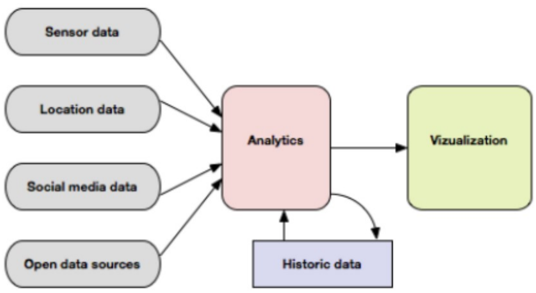

Figure 2. From data to decisions.

Thus, an analytical framework outlining the possible data sources, analytical and visualization techniques could be a valuable contribution to decision-making, as well as for future studies in this domain. Our research question for this study is "*Which data sources are applicable to the different application areas of smart cities*?" and "*which analytical techniques are available for analysis?"*

Data collection was done through several iterations. Initially, we planned on conducting this study as a pure literature review. However, there are few studies in this area so far. Using the Norwegian research library Oria (providing access to EBSCO, IEEE, JSTOR, PROQUEST and SAGE), we were only able to identify nine research papers (referenced in Table I) using the search phrases "smart city" and "decision-making" in the title. Thus, we had to rely on additional data sources and conduct a document analysis of industry white papers, as well as industry, technology and regular news sources. In addition, we applied existing empirical data from a previous study on air quality monitoring.

This exploratory approach led us to three themes, which we summarize in Section III, Table I. Further, the examined news sources and white papers identified nine application areas of data analysis in smart cities; parking, speed monitoring, public transport, traffic, environmental monitoring, energy management, waste handling, crime prevention, and home healthcare.

We conducted a second literature review round where we examined analytical techniques (intelligence layer). There were few, if any, studies explicitly combining smart cities and in-depth description of analytical technique, so we expanded our review and came up with 26 articles describing analytical techniques and methods relevant for smart cities. There were a lot more articles available, but the 26 we selected provides an overview of the most common methods and techniques. Snowballing from the reference lists of the articles revealed additional relevant references. We applied combinations of the following search phrases and keywords for the second round: "Big data analysis, tools, "research methods", statistics, "data analytics", "spatial data". In addition to the research papers, we have also examined additional web sources (digital methods initiative, Github). In both rounds of the literature review, we read the abstract and conclusions of the papers in order to identify which papers were relevant. The number of papers reported is what we were left with after this process.

For analysis, we have applied literature, findings from the air quality monitoring study, as well as data from industry to map potential data sources for each of the nine categories. This allowed us to create an initial framework of data sources for the nine identified categories.

III. Data for Decision-making

At the instrumentation layer, data for decision-making may originate from many different sources. Laney [12] defines big data as data having high volume, high velocity and/or high variety. High volume refers to large amounts of data-demanding both specialized storage and processing. High velocity refers to streams of real-time data, e.g., from sensor networks or large-scale transaction systems. Finally, high variety is about dealing with data from different sources having different formats.

Big data may originate from sensors. Another important source for big data is the world-wide-web. Web mining can be used to retrieve unstructured data (text) related to everyday events happening in a city. In this context social media, such as Facebook and Twitter can provide information about problems and citizen sentiments. Many government organization and private companies offer open data sets online that can be used for analysis and decision- making.

Marr [13] argues that the real value of big data is not in the large volumes of data itself, but in the ability to analyze vast and complex data sets beyond anything we could ever do before. Due to recent advances in data analysis methods and cloud computing, the threshold for using big data has diminished.

A. Sensors

Sensors and sensor networks are essential for smart decision-making. Sensors provide real-time information on a wide range of areas, such as weather, traffic, air quality, energy consumption, water consumption, and waste. Data from sensor networks are structured and easy to process, although different vendors and makers can introduce some difficulty. According to Cambridge dictionary, the word "sensor" means a device that is used to record that something is present or that there are changes in something. IoT is an infrastructure with interconnected units that may among other things act as sensor platforms. Botterman [14] defines IoT as:

"A global network infrastructure, linking physical and virtual objects, through the exploitation of data capture and communication capabilities. This infrastructure includes existing and evolving Internet and network developments. It will offer specific object-identification, sensor and connection capability as the basis for the development of independent federated services and applications. These will be characterized by a high degree of autonomous data capture, event transfer, network connectivity and interoperability". (p.12).

B. Location data

Location data places an object in a specific position. Location is important both for stationary and mobile objects. For mobile objects, location data comes from the Global Positioning System (GPS) or from triangulation of radio signals, e.g., belonging to a mobile network.

C. Social media

Another possible data source for smart decision-making is social media. Social media has been defined differently among scholars [15]. However, we adopt the definition by Kaplan and Haenlein [16]: "Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (p.62).

Data retrieved from social media will mostly be unstructured (text, images, video), but also structured meta-data providing additional information, e.g., tags containing author, content type, title, date/time and location. Unstructured data from social media may provide insight into the perceptions and sentiments of smart city citizens.

D. Open Data Sources

Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike. Open data has the following characteristics [17]:

- Availability and access: The data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading from the Internet. The data must also be available in a convenient and modifiable form.
- Reuse and redistribution: The data must be provided under terms that permit reuse and redistribution.
- Universal participation: Everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups.
- Interoperability: The ability to interoperate - or intermix - different datasets (i.e., one piece of open material contained therein can be freely intermixed with other open materials).

E. Decision-making in Smart Cities

In the context of smart city, decision-making has been given less attention in the literature; Google Scholar found nine articles discussing decision-making in smart cities (See Table I). The nine articles investigated various aspects of the three layers described earlier.

Studies related to the interconnection layer have highlighted various collaboration aspects that are important for smart cities. Ojasalo and Tähtinen [18] proposed a model of an open innovation platform for public sector decision-making in a city. The authors identified three different kinds of relationships that are present and partly interwoven in open innovation platforms (i.e., governing, sparring, and collaboration). The proposed model helps in organizing the three types of relationships of an innovation platform with the city's decision-making and external actors, by combining different decision-making cultures between the public and private sector.

TABLE I. MAPPING LITERATURE TO SMART CITY LAYERS

Ref.	Instrumentation layer	Interconnection layer	Intelligence layer	Others
[18]		X		
[19]			X	
[20]			X	
[21]			X	
[22]	X			
[23]	X			
[24]	X	X	X	
[25]				X
[26]				X

At the intelligence layer, Eräranta and Staffans [19] discussed knowledge creation and situation awareness in collaborative urban planning practice, and how digitalization changes it. The authors argued that smart city planning is not only a data-driven superlinear scaling practice, but an integrative and collaborative learning process facilitated by face-to-face interaction, advanced analyses and visualizations of available data, ongoing processes, and local history and stories. The authors brought in collaboration at the intelligence layer.

At the intelligence layer, Passe et al. [20] attempted to understand human behavior and decision-making about the built environment within an expanding range of spatial, political, and cultural contexts. The authors emphasized the importance of participation by a broad range of stakeholders in making decisions for the future of smart cities. The authors argued for the need to consider social dynamics in addition to building-occupant interactions, which requires investigating multiple scales and types of data to create new methodologies for design and decision-making processes. This approach moves data collection, analysis, design, and decision-making away from hierarchical relationships and utilizes the expertise of all stakeholders.

Also at the intelligence layer, Honarvar and Sami [21] talked about the various sensors embedded in different places of smart cities to monitor and collect data about the status of cities. Mining such data to extract valuable knowledge creates a challenge because various sources of data in smart cities are big, independent, heterogeneous and no semantic is integrated and annotated to them. The authors proposed an approach to leverage linked open data and semantic web technologies, data mining mechanisms, and big data processing platforms.

At the instrumentation layer, Khan et al. [22] emphasized the role of citizen participation as an important data source for social innovation and co-creating urban regeneration proposals through innovative IT systems. Those IT systems can use open government data, visualize urban proposals in 3D models and provide automated feedback on the feasibility of the proposals. Using those IT systems as a communication platform between citizens and city administrations offers an integrated top-down and bottom-up urban planning and decision-making approach to smart cities. In the same line, Foucault and Moulier-Boutang [23] proposed a governance model called “Smart City – organological”. The model consists of an adaptive device built around differentiation of smart sensors and tags to improve human decision-making. The device is taking into account both “physical sensors” and “economic and social sensors” to capture the explicit or implicit needs.

At the level of the three layers, Nathali Silva et al. [24] expressed concerns about the continuous growth of the complex urban networks that is challenged by real-time data processing and intelligent decision-making capabilities. The authors proposed a smart city framework based on big data analytics. The framework operates on three levels: instrumentation layer (data generation and acquisition level), interconnection layer (collecting heterogeneous data related to city operations, data management and processing level), and intelligence layer (filtering, analyzing, and storing data to make decisions and events autonomously, and initiating execution of the events corresponding to the received decisions).

Some other topics were studied in the literature, e.g., Gang and Yang [25] studied design issues to improve the intelligence layer of city emergency management. Kurniawan et al. [26] investigated the development and optimization possibilities of Makassar City smart operation room. The authors used fuzzy multi-criteria decision-making to illustrate the project priority rank and further to determine the alternative optimal option in conducting the project.

IV. Application Areas

In order to understand more about data sources and decision-making techniques, we have examined some common application areas connected to smart cities which were identified in the literature review (See Table II). The first four areas are connected to transport:

- Smart parking
- Speed monitoring
- Smart public transport
- Smart traffic

The rest of the application areas represent the broadness of the smart city concept:

- Environmental monitoring
- Energy management
- Waste handling
- Crime prevention
- Home healthcare

A. Smart Parking

Smart parking assists drivers to find a nearby parking spot. The information provided to the driver can have many different forms, from public displays placed next to roads to mobile apps directing the driver to a free parking spot [27][28][29].

Smart parking data is sensor based. Outdoor sensors may be magnetic sensors located in capsules embedded in the ground, detecting the presence of a car, or cameras detecting if a parking spot if free or not. Indoor parking spots may instead have infrared or ultrasound sensors to detect the presence of cars.

Smart parking may also include payment solutions based on mobile phone apps, use of SMS, or dedicated devices like SmartPark™ [30]. The payment solutions may give the user the opportunity to pay for time actually used instead for paying for a fixed time period.

Smart parking sensor data provides information to city planners and car park companies about the occupancy of parking spots over time. The collected information can be used for decision-making regarding the construction of new parking sites, and to decide on pricing.

B. Vehicle Speed Monitoring

Vehicle speed monitoring warns drivers about their driving speed. The idea is to make drivers slow down if they are driving at excess speed. Speed monitoring units may be stand-alone, but state-of-the-art units are connected to the Internet and provide real-time information on driving habits [31].

Several technologies have been demonstrated for vehicle speed monitoring including the use of cameras, RADAR, LIDAR, and underground sensors [32]. A measurement station is put in a fixed position, and excess speed is shown on a display device.

Another approach is to install mandatory units in all vehicles. The driver can then be alerted of excess speed directly by the unit. Such units can also upload speed data through some kind of network [32].

(Some GPS devices warn the driver about excess speed, but such data are not relevant, since data are not uploaded for use by traffic authorities.)

Vehicle speed monitoring data can be used by traffic authorities and police to decide on traffic control locations. Such data can also be used to implement speed reducing measures, such as speed bumps or traffic lights, and even control such measures in day-to-day operations.

C. Smart Public Transport

One essential measure to reduce environmental footprint is to reduce car traffic, in particular the use of private cars. Well-developed public transport infrastructure can be an incentive to reduce traffic load. Car owners may also be discouraged by the toll charges or congestion charges implemented in many cities.

Smart public transport uses technology to provide public transport users with a better user experience [33]. Use of sensors and GPS technology can provide real-time data on arrivals and departures of public transport vehicles.

Smart ticketing solutions may use smart cards or mobile phones equipped with Near Field Communication (NFC) to make ticketing more efficient from a user point of view [34].

Online route planners may help users choose the most efficient route from one location to another location.

The data collected from smart public transport can be used for real-time situation reports and may also be used by public transport planners to adjust timetables, change routes, create new routes, and adjust fares.

Social media may be mined to find citizen perceptions of the public transport system.

D. Smart Traffic

Smart traffic is about using technology to ensure more efficient traffic management. Traffic management may use road lights and signs to optimize traffic flow in real time [35]. Commercial car navigation systems provide information on fastest and shortest routes. Some navigation systems collect information from other cars real time to detect bottlenecks and provide alternative routes.

Data may come from sensors embedded in the roads. The most common technique is to detect traffic density by embedding coils under the road surface to pick up passing cars. Alternatives are to use cameras or radar technology to detect traffic.

Data may also come from the vehicles themselves, by using radio transmissions or a cellular network [36].

The data collected may be used by the city-administration for road-planning, adjusting intervals of traffic lights. Data can also be used by transport companies to decide on best schedules for pick-ups and deliveries.

Mining social media may provide some information on how citizens experience traffic situation.

E. Air Quality Monitoring

Monitoring air quality and other environmental parameters are the important for decision-making. Some cities are enforcing restrictions on traffic when pollution levels reach a certain threshold [37].

In most cases, the air quality monitoring is done by fixed monitoring stations located throughout the city, but may also be done by mobile handheld units, or units installed in cars.

Measurements include gases: CO, CO_2, NO_X, and dust particles, normally 2,5 PM and 10 PM.

Collected data can be combined with other data sources, e.g., meteorological data, to provide real-time situation reports and make forecasts for future pollution levels. Data can be visualized and be made available to the public. Such data is particularly valuable for citizens with respiratory problems.

Social media may be mined to find citizen perceptions of air quality.

F. Energy Management

Smart power grids contribute to better energy management and reduced environmental footprint. An essential part of the smart grid is smart meters. Smart meters are devices that continuously measure the power consumption of households and buildings. Household appliances can communicate with the smart meter to schedule activities when the load on the power grid is low. The smart meters also communicate with energy management systems to optimize energy consumption [38]. Buildings can also take part in energy production through the use of solar panels and other alternative energy sources.

Sensor data may be combined with location data and open data sources to make forecasts. Social media data plays a minor role in the context of energy management.

G. Waste Handling

Sorting waste materials for recycling has become common practice. Garbage collection can be improved by only collecting waste when necessary. “Intelligent” waste containers can report their state of becoming full and get included in the schedule of trucks collecting the waste [39][40][41].

The recycling process itself can provide valuable data on types and amounts.

Data from the waste collection process can be used to decide on container size and pick-up patterns. Data may also be made public to show timeliness and efficiency of the waste handling, from garbage collection through recycling.

Social media data mining can be used to detect sentiments about garbage collection.

H. Crime Prevention

Crime prevention is about allocating police resources to areas most likely to get victims of crime, but also to find out where to establish surveillance by video cameras and other means. Home or business security systems may discourage criminals and prevent crimes from being committed.

Data used for crime prevention will mostly be formerly reported crimes combined with open data sources, e.g., demographic data, property values, income levels of citizens, street light coverage, etc. [42][43].

Social media may be mined to find indications of unreported crimes.

I. Home Healthcare

Home health care is an important measure to enable healthcare patients to live in their homes as an alternative to nursing homes, and thereby reducing the burden on the healthcare system. Technology is important for patients to feel safe and to manage their health conditions. Safety alarms alert healthcare personnel about emergencies including fall detection. The safety alarm may have a built-in GPS device to provide healthcare personnel with the current location of the patient. Vanus et al. [44] describe how different sensors can be used to detect daily living activities in smart home care. Mshali, Lemlouma, Moloney, and Magoni [45] made a survey on health monitoring systems for use in homes. Smart medicine dispensers can alert the patient to take medication, and also notify healthcare personnel that medicine has been retrieved from the dispenser [46]. Sensor platforms are also used to monitor chronic diseases to make sure patients receive proper care [47]. Sensors may detect changes in medical conditions before the patients become aware of the change themselves [48]. Such data are important to make home healthcare smarter.

Open data may be used for planning purposes, e.g., statistics about the demography and increase of certain medical conditions. Social media does not play any significant role in home healthcare.

Table II summarizes the data sources applied in the different studies mentioned above.

TABLE II. MAPPING APPLICATIONS TO DATA SOURCES

Application areas	Data sources			
	Sensor data	Location data	Open data	Social media data
Smart parking	X	X	-	-
Speed monitoring	X	x	-	-
Smart public transport	X	X	-	x
Smart traffic	X	x	-	x
Air quality monitoring	X	x	X	x
Energy management	x	x	x	-
Waste handling	X	X	-	x
Crime prevention	-	X	X	x
Home health care	X	x	x	-

X major data source
x minor data source
- not applicable

V. Analytical technologies, methods and techniques

There are few research articles on big data in smart cities that are explicit about the actual methods, technologies and techniques being used for data analysis, so we had to rely on more general themed big data analytics articles. In this section, we provide an overview of these and attempt to link methods with application areas. In an overview section such as this, there is only room for a brief overview of each individual technology and method. For in-depth descriptions, we refer to the individual articles and papers referenced.

A. Technologies for storage and retrieval

Our literature review of big data analytical tools returned a lot of hits covering not so much methods as technologies. The reason for this is that big data requires somewhat different approaches in terms of storage and retrieval. Large amounts of data and the need for effective and selective filtering and retrieval are some of the challenges these technologies attempt to address [49]. For smaller data sets traditional techniques, or combinations of new and old, are just as appropriate. For example, software such as MS Excel can handle sheets of up to 1 million rows [50], and SPSS remains a powerful tool for statistical analysis.

Key technologies for big data include cloud computing, distributed file systems and distributed programming environments, as well as new database technologies such as NoSQL and NewSQL [51]. Cloud computing and associated technologies have the advantage of being scalable and flexible, and No/NewSQL databases have more flexible data structures and rules compared to traditional SQL databases, allowing for easier filtering, storage and retrieval of data [52]. As big data is often un- or semi-structured, there is a need for technologies that allow for redundancy and novel combinations of data for exploration and analytical purposes [53].

A typical package for big data analysis would include Apache Hadoop, using a file system such as HDFS for distributed storage, combined with a NoSQL database [54]; [55][56]. NoSQL databases can further be divided into (at least) four categories [57]:

- *Key-value*, where each value corresponds to a primary key, offering a simple and powerful structure for handling big data.
- *Bigtable* or *wide-column*, a more structured approach capable of handling large and complex data sets.
- *Document,* where entries as stored as documents rather than relations, and
- *Graph,* which stores information about nodes and their relations.

Each category has different use cases, which are beyond the scope of this section. Details can be found in the following referenced papers: Apache HBase and Cassandra are a few examples of popular Wide-column databases [58][59]. Oracle Berkeley [60] is a popular key-value database, where a previous iteration was involved in the first Bitcoin implementation. Mongo DB is a popular document database, where data is stored in documents rather than records, allowing for easily changing the data structure and varying the fields recorded for each piece of data.. Users include Amazon and Adobe [61]. Finally, Neo4J is an example of a graph database, used for areas such as bioinformatics, recommender systems and network graphs [62].

Distributed storage of data introduces some challenges related to processing and retrieval of data, which is where the map/reduce paradigm comes into play [63]. In short, map/reduce provides a programming model which creates a set of key-value pairs between records, regardless of where they are located in the distributed file system.

When the core technologies (file system, storage, database) are in place, the next layer is the analytical methods and techniques applied to analyze data.

B. *Analytical methods and techniques*

The literature review shows that there are many analytical methods in use, ranging from traditional statistical analysis to a plethora of machine learning methods and algorithms. There are also a few papers with a more critical perspective, warning researchers to not become deterministic and blind for the social construction of algorithms, but rather pay close attention to ethics and approach data-driven methods with a critical attitude and a thorough understanding of both domain and context [64]. An on-going survey from the Universities in Nottingham, Oxford and Edinburgh [65] is currently examining this, presenting respondents with case studies showing the outcome of different algorithms, and asking them to reflect on these different outcomes. We suggest that researchers from both academia and the private sector apply the same ethical and critical perspective to big data analysis as they would any other research project, and critically examine the fit between research question/hypothesis and the method being applied.

1) *Statistical methods*

Several papers combine Big data platforms with traditional statistical analysis. A study of terrorist ideology and attack type used a Hadoop platform to combine the global terrorism big data set with Google news data about terrorist attacks. The data was analyzed in SPSS using descriptive statistics and correspondence analysis [66]. A similar approach was applied to a study of electric vehicle customers [67]. However, for data sets that are "true" big data, with millions of cases, traditional statistical methods such as correlation will often show significant results between variables even where there is no real-life correlation, due to the sheer volume of data [68].

While traditional statistical methods have been successfully applied, several scholars point out that big data is often recognized by being unstructured and difficult to organize in variables such as we are used to from traditional statistical methods. Thus, new methods have emerged, such as sentiment, network, various link and content-based analytical techniques, and of course machine learning [69]. The remained of this section will provide an overview of some of the most commonly used methods. An excellent place to start could be the wiki of the *Digital methods initiative* [70], a collection of methods for digital research run by a consortium led by the University of Amsterdam.

2) *Text analytics and Sentiment analysis*

Text analytics, or text mining, refers to a set of methods for extracting and analyzing text-based content from news media, social networks, e-mails, blogs etc. [69]. One example using a set of text analytics methods is a study of hotel guest experiences, where guest ratings were analyzed to extract factors that were of particular importance to hotel guests [71]. Methods for text analytics include traditional content analysis [72] or various forms of discourse analysis [73], counting techniques such as word frequency, word count, word clouds [71]. Sentiment analysis has become a popular method in many areas such as politics [74][75], business and marketing [76]. Sentiment analysis is a classification process where words and phrases are classified as positive or negative in order to analyze public opinion on various things such as brands, political parties or current issues. Sentiment can be classified using pre-coded lists of words and phrases or as supervised machine learning [77].

3) *Social network analysis*

It is said that the Internet and social media has contributed to our current age being called the network society, as more and more of our lives can be seen as parts of a network [78]. Our Facebook and Twitter friends, LinkedIn contacts and the websites we follow, connect ideas and opinions. Social network analysis examines how information flows through a complex network and allows us to visualize and analyze networks by examining the connections and attributes of connections between nodes [79]. The basic use of network analysis is to identify patterns of interaction among the participants in a network. Typical variables measured are:

Degree: The number of participants a given participant interacts with, can be split into receiving (in-degree) and sending (out-degree) messages. High degree levels indicate strong networks and community. *Centrality*: How important a participant is to the network. Measured as closeness (the number of nodes between two participants), betweenness (how each participant helps connect other participants), and eigenvector (how well a participant is connected to other active participants). *Clustering*: The degree to which a set of participants form a group within the network. *Density*: The proportion of actual vs. potential connections within the network [80].

Social network analysis can for example be applied to understand how information flows between actors, such as in the study of disaster management after the Louisiana floods [81], or combined with graph databases when creating advanced recommender systems [82].

4) *Spatial analysis*

Combining data and location is a powerful analytical tool that has been around for a long time. In Utah, data from health records and environmental records were combined and used to predict areas likely to see more cases of cancer [83]. A similar study from Scotland combined geographical and demographic data to examine mortality rates in different parts of the country [84].

These and a range of similar studies rely on traditional compiled registry data, which by itself is a powerful tool. However, adding data sources such as sensors, data mining of the Internet etc can improve the predictive power of these models, for example in fields such as epidemiology, transportation, flooding and environment/climate studies. Using sensors from cars to collect data along with traditional registry data such as traffic congestion statistics, researchers were able to create a detailed spatial distribution of Carbon emissions in China [85]. Another study combined geographical data with accident statistics, data from taxis, public transport, and social media to create a predictive model of traffic accident hotspots [86].

5) Machine learning

Artificial intelligence and machine learning are some of the most talked about issues in current science. At its core, machine learning involves creating software that improves through experience, by being shown some examples which are run through various algorithms, allowing the software to learn "by itself" [87]. Machine learning is the preferred method for development of software for computer vision, speech recognition, natural language processing, robot control and more, and has also become prevalent in sciences ranging from biology to the social sciences [88], as big data is making traditional methods more difficult.

At its core, machine learning consists of feeding data to a piece of software and applying various algorithms (step-by-step instructions) to make sense of the data [89]. There are hundreds of techniques and even more algorithms for machine learning, but all the different techniques can be categorized into two categories: Supervised and unsupervised [89]. A review of machine learning literature found 156 supervised and 46 unsupervised algorithms tested in 121 different studies [89].

Supervised machine learning involves feeding data to an algorithm along with a set of labeled "training data" [90]. For example, sentiment analysis could be conducted by manually coding a data set and feeding this to an algorithm which would then use the example data to continue coding more data. In order to verify the model, several iterations of evaluation are needed to calculate accuracy [90].

Supervised learning techniques include the following:

- *Regression*: for calculating the relationships between variables (several algorithms for big data regression).
- *Classification trees:* Where variables are split into multiple dimensions to form a tree structure.
- *Ensemble learning/aggregation*: Additional training techniques to ensure the accuracy of tree models.
- *Support vector machines*: Another method of classification.
- *Neural networks*: Often used for learning purposes such as speech recognition or social networks, Neural networks are made up of nodes and connections receiving and sending signals. Originally modeled on the human brain, but has since branched out in many directions.
- *Nearest neighbors*: Used in pattern recognition, regression and classification. For example, given the length of petals for a set of flowers, the algorithm can identify the different types of flower [89].

Unsupervised machine learning on the other hand, does not involve any manual coding. Data is fed unlabeled to the algorithm in order to make sense of it [89]. Some of the popular techniques include *Clustering*, where the purpose is to find connections between variables. Clustering is a popular technique in marketing and recommender systems, as it can discover what people with a certain set of characteristics typically buy or are interested in. Combining demographic data, purchase data, interests, or text from the news articles we read are typical data sources [91]. Another popular technique is *dimensionality reduction*, where the objective is to reduce the number of variables under consideration. For example, sensor-based location data can come from a number of different sources, with each source providing overlapping information. Dimensionality reduction can help reduce the number of variables into one "position" variable [92].

The methods reviewed in this section are summarized in Table III.

TABLE III. ANALYTICAL METHODS AND APPLICATIONS

Method	Techniques/application	
Social network analysis	Identify attributes, relations and networks (of networks)	
Sentiment analysis	Identify negative/positive opinions related to a topic/case/issue	
Spatial data analysis	Combine data and location to visualize where something happens/could happen	
Statistical analysis	Find relations between variables	
Machine learning	Unsupervised techniques	Find connected variables. Reduce dimensions (variables) providing data on the same thing.
	Supervised techniques	Classification and regression (relations) between different variables. Pattern recognition.

VI. ANALYTICAL FRAMEWORK

The purpose of our case studies is to examine data sources and methods used to analyze data. Seven examples of smart city applications show the importance of sensor data, but also the opportunities for using open data sets combined with sensor data to improve analysis and enable forecasting. Web mining and social media have limited use in these cases, but can be used to alert city administration about potential problems and sentiments.

The crime prevention case does not rely on sensor data, but on reports of crimes. Combining different open data sets can provide better insight related to crime prevention. The reported crimes can provide patterns, but combining data sets may shed light on underlying factors, like property values, incomes, absence of street lights and other factors.

In this study, we have examined mainly the instrumentation and interconnection layers, finding a set of data sources used in different smart city application areas, as shown in Table II.

When we map these findings to the three layers in Table I, we have the outline of an analytical framework as shown in Figure 3. The resulting analytical framework may guide future research efforts in the field.

Existing research and white papers provide examples of how big data can be applied for decision making, but as our framework shows, there is a need for both synthesizing existing studies as well as conducting new empirical studies to create a roadmap for decision-makers.

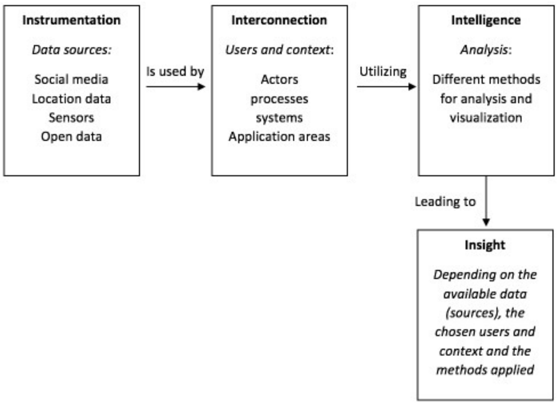

Figure 3. Analytical framework.

This roadmap would list relevant data sources and analytical techniques for different users and contexts. The framework forms a possible foundation for future studies in this area.

VII. Conclusion, Limitations and Future Work

In this article, we used nine common application areas of smart cities to explore their use of data. We examined relevant data sources, and their use. Data collected from sensors are very important for seven of the chosen application areas. Open data is often a valuable supplement to collected data. In some cases, location data are combined with other types of data. Social media data mining may play a role to show user perceptions and sentiments.

The collected data need to be processed and analyzed to be useful for decision-making.

Data will often be used for automatic decision-making. In eight of the chosen application areas, we found examples of data used for automatic decision-making:

- Smart parking: Automatic update of displays directing drivers to available parking spots.
- Speed monitoring: Automatic regulation of traffic lights, or even photographing speeding vehicle to issue a speed ticket.
- Smart public transport: Automatic updates of screens showing arrival and departure times.
- Smart traffic: Automatic control of signs and traffic lights to redirect traffic.
- Air quality monitoring: Automatic alerts to citizens in the areas, through signs or SMS service.
- Energy management: Automatic start of household appliances, based on grid load.
- Waste handling: Automatic updates of garbage truck schedules based on amount of garbage in each container.
- Home healthcare: Automatic requests for health care personnel to look into changing medical situation for patients living in their own homes.

For strategic and long-term decisions done by humans, the results of the processing and analysis need to be visualized in a meaningful way, e.g., through graphs, bar charts, pie charts often combined with a map or even embedded in a Geographic Information System (GIS) front-end.

However, researchers analysts should be aware that algorithms and methods are social constructions, and that the choice of algorithm and method in some cases will influence the outcome of research. Careful ethical and methodological considerations should therefore be considered when planning and implementing an analytics project.

This article studied application areas of smart cities and methods for data analysis to examine use of (big) data. The study is not exhaustive. We used example of application areas from literature, but as "smart cities" have ambiguous definitions, we may have overlooked some areas. Further, as we had to start examining white papers from industry it is likely we have missed interesting data from relevant sources even after our rigorous search in the most well-known big data/analytics companies.

A. Future research challenges

Finally, we would like to present some future research challenges derived from this literature review of smart city and data analytics.

Consolidation: Data analytics and smart cities are both new and intertwined fields, with research coming from both highly technical and (some) organizational fields. There are few studies that combine a technical solution with a thorough field test and case study. Thus, we argue there is a need for consolidation of the field, in order to move it towards the mainstream. City managers need off the shelf tools and simple processes in order to make use of analytics.

Analytical methods: We have attempted to summarize the analytical methods identified in literature in Table III. However, further work is needed in order to classify these methods for managers and the organizational level. Most of the literature on these methods were written for a technical audience, requiring skills that decision-makers may not have.

Application areas: The list of smart city application areas was selected based on the literature review. There are many other application areas, but the selected areas were considered to provide good examples on both data sources and analytical methods. Future work may include even more application areas.

Actors and process: In our framework (Figure 3) we include actors and process (including context) as factors. Few of the studies included in this review address these in detail, except brief mentions of the actors involved. As context is important in technological implementations, there is a likely need for more in-depth studies of how process and actors/stakeholders influence/are influenced by, the analytical process.

Finally, we intend to investigate further the use cases for smart cities' use of analytics and big data, so that we can present a comprehensive model of possible combinations of data sources, actors and contexts, and analytical techniques.

For practitioners and researchers with little technical background, a handbook of methods, techniques and use cases would be a valuable tool that could potentially improve and simplify data analysis strategies and outcomes.

References

[1] L. Berntzen, M. Rohde Johannessen and R. El-Gazzar, "Smart Cities, Big Data and Smart Decision-making - Understanding ´Big Data´ in Smart City Applications", Proceedings of ICDS 2018.

[2] ITU-T. *Focus Group on Smart Sustainable Cities.* [Online]. Available from: https://www.itu.int/en/ITU-T/focusgroups/ssc/Pages/default.aspx [retrieved: 2018.02.04].

[3] ITU-T Focus Group on Smart Sustainable Cities, "*Smart sustainable cities: An analysis of definitions*. [Online]. Available from: https://www.itu.int/en/ITU-T/focusgroups/ssc/Documents/website/web-fg-ssc-0100-r9-definitions_technical_report.docx, 2014. [retrieved: 2018.02.04].

[4] A. Vesco and F. Ferrero, Eds., Handbook of Research on Social, Economic, and Environmental Sustainability in the Development of Smart Cities. IGI Global, pp. xxv-xxxi, 2015.

[5] Deloitte. *Smart Cities: Big Data*. [Online]. Available from: https://www2.deloitte.com/content/dam/Deloitte/fpc/Documents/services/systemes-dinformation-et-technologie/deloitte_smart-cities-big-data_en_0115.pdf, 2015. [retrieved: 2018.02.04].

[6] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog Computing; A Platform for Internet of Things and Analytics," in Big Data and Internet of Things: A Roadmap for Smart Environments, N. Bessis and C. Dobre, Eds. Studies in Computational Intelligence, 546, Springer, pp. 169–186, 2014.

[7] V. Albino, U. Berardi, and R. M. Dangelico,. "Smart Cities: Definitions, Dimensions, Performance, and Initiatives," Journal of Urban Technology, 22(1), pp. 3–21, 2015.

[8] IBM. *A vision of smarter cities*. [Online]. Available from: https://www-03.ibm.com/press/attachments/IBV_Smarter_Cities_-_Final.pdf, 2010. [retrieved: 2018.02.04].

[9] M. Kehoe et al., Smarter Cities Series: A Foundation for Understanding the IBM Approach to Smarter Cities, IBM Redguides for Business Leaders, pp. 1–30, 2011.

[10] E. Negre and C. Rosenthal-Sabroux, "Smart Cities: A Salad Bowl of Citizens, ICT, and Environment," in Handbook of Research on Social, Economic, and Environmental Sustainability in the Development of Smart Cities, A. Vesco and F. Ferrero, Eds., IGI Global, pp. 61-78, 2015.

[11] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, 36(4), pp. 1165–1188, 2012.

[12] D. Laney, 3D Data Management: Controlling data, volume, velocity, and variety. Technical Report. META Group, 2001.

[13] B. Marr, Big Data – Using Smart Big Data Analytics and Metrics to Make Better Decisions and Improve Performance. John Wiley & Sons Ltd, 2015.

[14] M. Botterman, Internet of Things: An Early Reality of the Future Internet, Workshop Report, European Commission, Information Society and Media Directorate, 2009.

[15] J. W. Treem and P. M. Leonardi, 7 Social Media Use in Organizations Exploring the Affordances of Visibility, Editability, Persistence, and Association. Communication Yearbook, 36, pp. 143–189, 2012.

[16] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Business Horizons, 53(1), pp. 59–68, 2010.

[17] Open Knowledge International. *Open Data Handbook.* [Online] http://opendatahandbook.org/guide/en/what-is-open-data/ [retrieved: 2018.02.04].

[18] J. Ojasalo and L. Tähtinen, "Integrating Open Innovation Platforms in Public Sector Decision Making: Empirical Results from Smart City Research," Technology Innovation Management Review, 6(12), pp. 38–48, 2016.

[19] S. Eräranta and A. Staffans, "From Situation Awareness to Smart City Planning and Decision Making," Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM 2015), J. Ferreira and R. Goodspeed, Eds., Paper 197, pp. 1-17, 2015.

[20] U. Passe et al., "Methodologies for Studying Human-Microclimate Interactions for Resilient, Smart City Decision-Making, " Proceedings of the 32nd International Conference on Passive and Low Energy Architecture, P. La Roche and M. Schiler, Eds., pp. 1735-1742, 2016.

[21] A. R. Honarvar and A. Sami, "A Multi-source Big Data Analytic System in Smart City for Urban Planning and Decision Making," International Conference on Internet of Things and Big Data, Doctoral Consortium (DCIT), pp. 32-36, 2016.

[22] Z. Khan et al., "Developing Knowledge-Based Citizen Participation Platform to Support Smart City Decision Making: The Smarticipate Case Study," Information, 8, 47, pp. 1-24, 2017.

[23] J.-P. Foucault and Y. Moulier-Boutang, "Towards economic and social 'sensors': Condition and model of governance and decision-making for an organological Smart City," International Conference on Smart and Sustainable City and Big Data (ICSSC), pp. 106-112, 2015.

[24] B. Nathali Silva, M. Khan, and K. Han, Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making. Wireless Communications and Mobile Computing, pp. 1–12, 2017.

[25] L. I. Gang and L. I. Yang, "Construction of Emergency Decision-making Intelligence System Against the Background of Smart City," Journal of Library Science in China, 3(4), 2016.

[26] F. Kurniawan, A. P. Wibawa, Munir, S. M. S. Nugroho, and M. Hariadi, "Makassar Smart City Operation Center Priority Optimization using Fuzzy Multi-criteria Decision-making," 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), pp. 1-5, 2017.

[27] Smart Parking Ltd. *Company website*. [Online]. http://www.smartparking.com [retrieved: 2018.02.04].

[28] Fybr. *Company website*. [Online]. http://www.fybr.com [retrieved: 2018.02.04].

[29] WorldSensing. *Company website*. [Online]. https://www.worldsensing.com/industries/parking-operators/ [retrieved: 2018.02.04].

[30] SmartPark. *Company website*. [Online] https://smartpark.co.nz [retrieved: 2018.02.04].

[31] C. H. Schaffer, Customer Success Is Key – How a small manufacturer transformed an Internet of Things (IoT) solutions provider and unlocked $2 million is SaaS revenue. (Kindle edition) amazon.com, 2015. [retrieved: 2017.12.01].

[32] G. Kirankumar, J. Samsuresh, and G. Balaji, "Vehicle Speed Monitoring System [VSM] (Using RuBee Protocol)," IACSIT International Journal of Engineering and Technology, Vol. 4, No. 1, pp. 107-110, 2012.

[33] R. M. John et al., "Smart public transport system," International Conference on Embedded Systems (ICES), pp. 166-170, 2014.

[34] P. Chowdhury, P. Bala, and D. Addy, "RFID and Android based smart ticketing and destination announcement system," Advances in Computing Communications and Informatics (ICACCI), pp. 1-5, November 2016.

[35] R. Hawi, G. Okeyo, M. Kimwele, "Smart Traffic Light Control using Fuzzy Logic and Wireless Sensor Network," Computing Conference, London, pp. 450-460, 2017.

[36] K. Kumarmanas, S. Praveen, V. Neema, and S. Devendra, "An Innovative Device for Monitoring and Controlling Vehicular Movement in a Smart City," Symposium on Colossal Data Analysis and Networking (CDAN), pp. 1-3, 2016.

[37] A. Florea et al., "Low cost mobile embedded system for air quality monitoring - air quality real-time monitoring in order to preserve citizens' health," Sixth International Conference on Smart Cities, Systems, Devices and Technologies, (SMART), IARIA, pp. 5-12, 2017.

[38] C. Meinecke, Potentiale und Grenzen von Smart Metering. Springer. 2015.

[39] F. Folianto, Y. S. Low, W. L. Yeow, "Smartbin: Smart Waste Management System," IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1-2, 2015.

[40] A. S. Wiaya, Z. Zainuddin, and M. Niswar, "Design a smart waste bin for waste management," 5th International Conference on Instrumentation, Control, and Automation (ICA), pp. 62-66, 2017.

[41] H. Poddar, R. Paul, S. Mukherjee, B. Bhattacharyya, "Design Of Smart Bin For Smarter Cities," International Conference on Innovations in Power and Advanced Computer Technologies [i-PACT2017], IEEE Press, pp. 1-6, 2017.

[42] F. Wang, Ed., Geographic Information Systems and Crime Analysis, Idea Group Publishing, 2005.

[43] S. Chainey and J. Ratcliffe, GIS and Crime Mapping, Wiley, 2005.

[44] J. Vanus, J. Belesova, R. Martinek, J. Nedoma, M. Fajkus, P. Bilik and J. Zidek, "Monitoring of the daily living activities in smart home care," Human-centric Computing and Information Sciences, (7)30), pp. 1-34, 2017.

[45] H. Mshali, T. Lemlouma, M. Moloney and D. Magoni, "A survey on Health Monitoring Systems for Health Smart Homes," International Journal of Industrial Ergonomics, Elsevier (66), pp. 26-56, 2018.

[46] B. Kon, A. Lam and J. Chen, "Evolution of Smart Homes for the Elderly," Proceedings of the 2017 International World Web Conference, pp. 1095-1101, 2017.

[47] P. Leijdekkers, V. Gay and E. Lawrence, "Smart Homecare for Health Tele-monitoring," Proceedings of the First International Conference on the Digital Society, IARIA, 2007.

[48] S. Majumder, E. Aghayi, M. Noferesti, H. Memarzadeh-Tehran, T. Mondal, Z. Pang and M. J- Deen, "Smart Homes for Elderly Healthcare – Recent Advances and Research Challenges, Sensors, 17, 2496, 2017

[49] A. Katal, M. Wazid and R. Goudar, "Big data: issues, challenges, tools and good practices," Sixth International Conference on Contemporary Computing (IC3), pp. 404-409, 2013.

[50] P. Louridas and C. Ebert, "Embedded Analytics and Statistics for Big Data," IEEE Software, 30(6), pp. 33-39, 2013.

[51] J. Romero, S. Hallett, and S. Jude, "Leveraging Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector," Sustainability, 9(12), 2160, 2017.

[52] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics – Challenges in topic discovery, data collection, and data preparation," International Journal of Information Management," 39, pp. 156-168, 2018.

[53] Nosql-database.org, [Online] http://nosql-database.org/ [Retrieved 2018.09.01].

[54] A.S. Hashmi and T. Ahmad, "Big Data Mining: Tools & Algorithms. International Journal of Recent Contributions from Engineering, 4(1), pp. 36-40, 2016.

[55] F. Nasution, N. E. B. Bazin and A. Z. Daliyusmanto, "Big Data's Tools for Internet Data Analytics: Modelling of System Dynamics," International Journal on Advanced Science, Engineering and Information Technology 7(3), pp. 745-753, 2017.

[56] C. J. M. Tauro, S. Aravindh and A. B. Shreeharsha, "Comparative study of the new generation, agile, scalable, high performance NOSQL databases," International Journal of Computer Applications, 48(20), pp. 1-4, 2012.

[57] K. M. Anderson, A. A. Aydin, M. Barrenechea, A. Cardenas, M. Hakeem and S. Jambi, "Design Challenges/Solutions for Environments Supporting the Analysis of Social Media Data in Crisis Informatics Research," 48th Hawaii International Conference on System Sciences, pp. 163-172, 2015.

[58] Y. Huang, M. Lin and F. Yu, "Analysis on financial development correlativity between Zhejiang, Jiangsu, and Shanghai," WIT Transactions on Information and Communication Technologies 49, 2014.

[59] Oracle, "Oracle Berkeley," [Online] https://www.oracle.com/database/berkeley-db/db.html, [retrieved: 2018.09.01], 2018.

[60] K. Chodorow, "MongoDB: The Definitive Guide: Powerful and Scalable Data Storage, " O'Reilly Media Inc.", 2013.

[61] J. J. Miller, "Graph database applications and concepts with Neo4j," Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA, 2013

[62] A. Katal, M. Wazid, and R. H. Goudar. "Big data: issues, challenges, tools and good practices," Sixth International Conference on Contemporary Computing (IC3), IEEE, pp. 404–409, 2013.

[63] C. Fuchs, "From digital positivism and administrative big data analytics towards critical digital and social media research," European Journal of Communication, 32(1), pp. 37-49, 2017.

[64] Unbias Research Team, "Algorithmic Preference Survey," [Online] https://unbias.wp.horizon.ac.uk/2018/07/20/unbias-algorithmic-preference-survey/ [Retrieved 2018.09.01]

[65] K. D. Strang and Z. Sun, "Analyzing Relationships in Terrorism Big Data Using Hadoop and Statistics," Journal of Computer Information Systems, 57(1), pp. 67-75, 2017.

[66] R. G. Qiu, K. Wang, S. Li, J. Dong and M. Xie, "Big data technologies in support of real time capturing and understanding of electric vehicle customers dynamics," 5th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2014.

[67] S. T. McAbee, R. S. Landis and M. I. Burke, "Inductive reasoning: The promise of big data," Human Resource Management Review, 27(2), pp. 277-290, 2017.

[68] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management. 35(2), pp. 137-144., 2015.

[69] Digital Methods Initiative, Wiki [Online] https://wiki.digitalmethods.net/ [Retrieved 2018.09.01]

[70] Z. Xiang, Z. Schwartz, J. H. Gerdes Jr and M. Uysal, "What can big data and text analytics tell us about hotel guest experience and satisfaction?" International Journal of Hospitality Management, 44, 120-130, 2015.

[71] Ø. Sæbø, "Understanding Twitter Use among Parliament Representatives: A Genre Analysis," In E. Tambouris, A. Macintosh, & H. de Bruijn (Eds.), Electronic Participation (Vol. 6847, pp. 1-12): Springer Berlin / Heidelberg, 2011.

[72] M. Granath, "The Smart City – how smart can 'IT' be? : Discourses on digitalisation in policy and planning of urban development," Linköping University, [Online] http://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A956501&dswid=-3723 [Retrieved 2018.09.01]

[73] A. Al-Rawi, "Online political activism in Syria: Sentiment analysis of social media," SAGE Research Methods Cases, 2017.

[74] J. O. Øye, "Sentiment Analysis of Norwegian Twitter Messages. (Master Thesis)," NTNU, Trondheim, 2014.

[75] H. Shirdastian, M. Laroche and M. O. Richard, "Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter," International Journal of Information Management, 2017.

[76] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, 5(4), pp. 1093-1113, 2014.

[77] M. Castells, "The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance. The Annals of the American Academy of Political and Social Science, 616(1), pp. 78-93, 2008.

[78] J. Scott, "Social network analysis," Sage, 2017.

[79] A. Mazur, C. Doran and P. R. Doran, " The use of social network analysis software to analyze communication patterns and interaction in online collaborative environments," International Conference on Education, Training and Informatics, ICETI," Orlando, FL, 2010.

[80] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," International Journal of Information Management, 38(1), pp. 86-96, 2018.

[81] Y. Wang, L. Kung, W. Y. C. Wang and C. G. Cegielski "An integrated big data analytics-enabled transformation model: Application to health care," Information & Management, 55(1), 2017.

[82] W. Ball, S. LeFevre, L. Jarup and L. Beale, " Comparison of different methods for spatial analysis of cancer data in Utah," Environmental Health Perspectives, 116(8), pp. 1120-1124, 2008.

[83] S. Sridharan, H. Tunstall, R. Lawder, and R. Mitchell, "An exploratory spatial data analysis approach to understanding the relationship between deprivation and mortality in Scotland," Social Science & Medicine, 65(9), pp. 1942-1952, 2007.

[84] Z. Huang, F. Cao, C. Jin, Z. Yu and R. Huang, "Carbon emission flow from self-driving tours and its spatial relationship with scenic spots – A traffic-related big data method," Journal of Cleaner Production, 142, 946-955, 2017.

[85] K. Xie, K. Ozbay, A. Kurkcu and H. Yang, "Analysis of Traffic Crashes Involving Pedestrians Using Big Data: Investigation of Contributing Factors and Identification of Hotspots," Risk Analysis, 37(8), pp. 1459-1476, 2017

[86] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, 349(6245), pp. 255-260, 2015.

[87] P. Domingos, "A few useful things to know about machine learning," Communications of the ACM, 55(10), pp. 78-87, 2012.

[88] I. Portugal, P. Alencar and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," Expert Systems with Applications, 97, pp. 205-227, 2018

[89] C. Crisci, B. Ghattas and G. Perera, "A review of supervised machine learning algorithms and their applications to ecological data. (Report). Ecological Modelling, 240, 113, 2012.

[90] Z. Fan, S. Chen, L. Zha, L. and J. Yang, "A Text Clustering Approach of Chinese News Based on Neural Network Language Model," International Journal of Parallel Programming, 44(1), 198-206, 2016.

[91] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," The Journal of Machine Learning Research, 16(1), pp. 2859-2900, 2015