



# Big Data & Analytics for Societal Impact: Recent Research and Trends

Ashish Gupta<sup>1</sup> · Amit Deokar<sup>2</sup> · Lakshmi Iyer<sup>3</sup> · Ramesh Sharda<sup>4</sup> · Dave Schrader<sup>5</sup>

Published online: 18 March 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## 1 Introduction

Recent strides in computing capabilities, increases in data transparency and open data sources, growth in IOT (Internet of Things) and smartphone device usage are some of the drivers helping bridge the worlds of data, people and things. Increased availability of disparate and complex data requires the use of new analytics approaches. Such complex datasets or big data, defined in numerous domains, are traditionally characterized by the presence of large volume, high velocity, and high variety. Such data may be machine-generated or human-generated, depending on application contexts. Machine-generated data can have origins in autonomous systems such as robotic facilities in manufacturing systems, sensors such as RFID tags in supply chain touchpoints like warehouses and stores, environmental/weather related sensors, streaming sources like wearables, and so forth. Human-generated data may be generated through activities like social media engagement, mobile phone usage, forums on online platforms, search/purchase activities, and so forth.

Analytics applications have already started to make measurable impact on the organizations (Chen et al. 2012). Several recent editorials have focused on these issues. For example, Baesens et al. (2016) recognized the disruptive impact of big data and identified some of the innovative applications of big data such as online-to-offline commerce, proactive customer care, and IoT enabled cars. Pick et al. (2017) discussed location analytics for decision support in geospatial phenomenon (e.g., location-based services) that can have societal applications such as detecting patterns of diffusion of information, crime, disease and so forth. Chen et al. (2012) also discussed broad application areas such as science and technology, smart health and well-being, and security and public safety. In this paper, we highlight select big data applications such as lifestyle, disaster relief, energy and sustainability, critical infrastructure, and so forth that indicate promise for making a societal impact through the use of analytics.

We propose a simple framework to understand the research on big data applications for societal impact (Fig. 1). The three concentric circles represent the a) data and the infrastructure, b) techniques for big data analysis and interpretation, and c) application domains. The data and the infrastructure needed to collect, store and process the data form the fundamental building block for developing big data applications. It is represented by the innermost circle. Infrastructure includes the hardware and software needed for big data processing and could be housed locally in a data center or virtually in the cloud. Various cloud-based technologies such as Amazon Web Services (AWS), Google Cloud Platform, and Microsoft Azure that are used as virtual big data clusters have become more stable over years and are a viable option if establishing a localized data center is prohibitive for reasons such as cost, cluster maintenance, information security, and so forth. Various data repositories that could be utilized are traditional relational databases (e.g., Oracle, SQL Server, Teradata) or NoSQL (schema-less) databases such as columnar (e.g., HBase, Cassandra), as well as key-value based (e.g., Redis, Riak, MemcacheDB), document-based (e.g., MongoDB, CouchDB, Couchbase), and graph-based systems (e.g.,

✉ Ashish Gupta  
ashish.gupta@auburn.edu

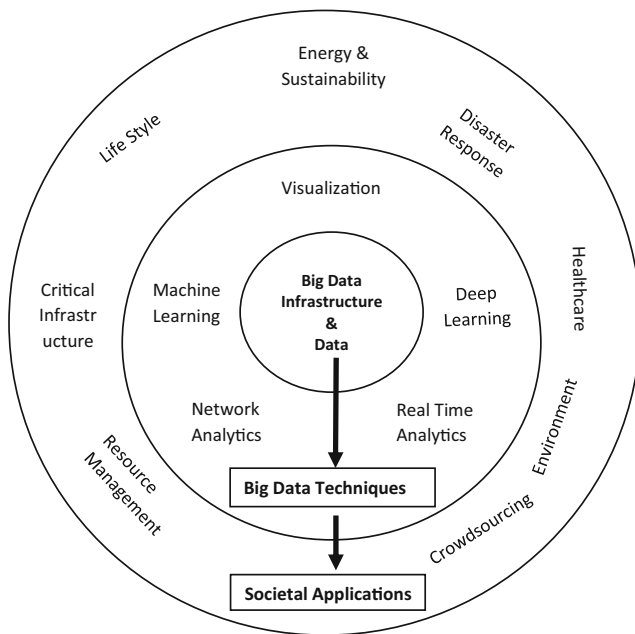
Amit Deokar  
amit\_deokar@uml.edu

Lakshmi Iyer  
iyerLs@appstate.edu

Ramesh Sharda  
Ramesh.sharda@okstate.edu

Dave Schrader  
drdaveschrader@gmail.com

- <sup>1</sup> Auburn University, Auburn, AL, USA
- <sup>2</sup> University of Massachusetts Lowell, Lowell, MA, USA
- <sup>3</sup> Appalachian State University, Boone, NC, USA
- <sup>4</sup> Oklahoma State University, Stillwater, OK, USA
- <sup>5</sup> Teradata University Network, Dayton, OH, USA



**Fig. 1** Framework for societal applications of big data

Neo4J, OrientDB). Environment such as Hadoop and Apache Spark typically are suitable for processing big data. While Hadoop utilizes Map and Reduce operations to speed up data processing using a divide-and-conquer approach, Spark uses RDD (Resilient Data Dataset) as a fundamentally different data structure abstraction for distributed collection of data objects that may be operated on different nodes of the cluster in a parallel manner to achieve better computational efficiency than MapReduce operations.

The middle layer highlights analytical approaches. Various machine learning, visualization and network analytics approaches can then be applied to make further sense of the data. Another emerging approach that is gaining tremendous traction among industry and scientific community is Deep Learning or Deep Neural Nets. Various types of Deep Learning approaches such as convolutional neural nets and recursive neural nets can outperform many traditional machine learning approaches. Big data environments identified in the innermost layer clearly impact the implementation and even the choice of analytical techniques that can be used for analysis. Finally, the outside circle represents the area of societal importance where various data analytic approaches could be applied to reap value. In the following sections, we describe selected areas of societal importance that benefit from big data applications.

## 2 Applications for Societal Impact

### 2.1 Healthcare

According to reports from the World Health Organization (WHO) and US Department of Health & Human Services

(HHS), healthcare spending comprises over 17% of US Gross Domestic Product. Despite this spending, the performance of US healthcare systems is not among the best. A combination of several factors such as fragmentation of care, lack of care coordination and continuity, poor lifestyle choices, lack of interoperable data and systems, etc., may be contributing to the inefficiencies and poor outcomes (Gupta and Sharda 2013; Gupta et al. 2013). With the increased use of Electronic Health Record (EHR) technologies, more patient and healthcare data is now available at hospitals that can be utilized for tasks beyond traditional medical billing applications. Data analytics can provide valuable insights when data from various points in time and space are integrated. For example, EHR data when combined with lifestyle data, care coordination data (e.g. continuity of care documents), claims data, and Patient Health Records (PHR) can be used to develop a patient's 360-degree view of health (Pickering et al. 2015). Illustrative areas that hold much promise for IS researchers to contribute to the future healthcare systems are social and network analytics, sensor (IOT driven) informatics for disease prevention, and imaging informatics for improved diagnostics and detection. We briefly introduce these topics and related papers in this special issue.

Network analytics can help understand the diffusion of certain disease patterns such as new pandemic spread. Social network approaches have demonstrated good results in applications such as influenza outbreak and smoking side effects. Applying network analytics approaches to the free flowing social media data helps segment the features and the nature of underlying networks, which can provide guidance in terms of not only clinical decision-making but also in effective resource-allocation and policy-making. Additionally, the healthcare industry generates continuous streams of data. For example, the data being recorded for each patient in a hospital at some frequency is a stream. Analysis of such data requires stream analytics. The specific patterns in the events can indicate some imminent outcomes such as the state of an organ like heart or lung. Kalgotra et al. (2017) present an example of how such data needs to be organized for facilitating analytics. They also illustrate the application of their data organization method by using data from an electronic medical record (EMR) system to show how health of tobacco users' progresses in certain diseases compared to non-tobacco users. Interestingly, some results are counter-intuitive; many diseases progress the same way in both tobacco users and non-users.

Use of sensors in healthcare has opened up a plethora of research opportunities for analytics applications in areas that have traditionally relied on only observational data that is typically limited (e.g., movement data) or self-reported data that may be biased (e.g. self-reported injuries). For example, Wilkerson et al. (2016) demonstrate the use of inertial movement units (IMU or sensors) in the sports context. They capture individual football player's data during the entire season

objectively and accurately capture their movement data using accelerometers and combine that data with the students' medical records to make recommendations for interventions to help prevent injuries. Such approaches could also be extended to hazardous occupational settings where improvements in the workflow processes or the addition of an integrative sensor-based mobile technology could be accomplished (Wilkerson et al. 2018) that drives making recommendations or detecting performance patterns.

Finally, imaging recognition provides opportunities for IS researchers to explore the next generation of analytics approaches such as deep learning and reinforcement learning. Recent improvements in the processing capabilities of graphical processing units (GPU) and in-memory capabilities is contributing to the huge growth in revolutionary approaches such as deep learning to gain improved specificity and sensitivity along with automatic feature extraction. Various types of deep neural nets can be used depending on the task to be accomplished. For example, recurrent neural nets could be used for analyzing streams of data such as social media data while convolutional neural nets can be effectively utilized to process 2D and 3D images (Ravi et al. 2017). Gruetzemacher & Gupta (2016) demonstrate how a large database of annotated CT scans was used to train deep nets for cancer nodule detection. Further research could be done on functional magnetic resonance imaging (fMRI) for neuroscience applications, larger cancer data sets, as well as applications in the marketing domain that rely on image sharing.

## 2.2 Lifestyle

Healthcare focus, especially in North America, is shifting from being reactive to proactive. For example, much of the previous healthcare analytics research has mainly looked at claims data to assist healthcare providers in developing a better understanding of the services rendered. A Robert Wood Johnson Foundation (2014) study on health IT shows that 80% of health outcomes are associated with factors that are outside of the traditional health services, such as behavioral and economic factors. Given that only 20% of health outcomes can be managed through clinical services, the focus in health management is shifting to improving overall lifestyle of an individual rather than just when he/she is a patient (Bresnick 2016). Combining lifestyle data with EHR data can allow providers to push appropriate lifestyle management strategies to consumers for improving their personal health and thereby mitigate hospitalization. Fitzgerald (2015) highlights Intermountain Healthcare as a pioneer in its efforts to use both lifestyle and EHR data to “discover the power of data analytics to affect population health”.

The rapid explosion of mobile and evolving big data technologies offer huge potential to empower individuals to take ownership of their personal well-being through proper

lifestyle management. In addition, as explained earlier, wearable sensors for both medical and lifestyle purposes have increased the volume, velocity, and variety of data providing the opportunity for analytics to improve population health (Kankanhalli et al. 2016). Andreu-Perez et al. (Andreu-Perez et al. 2015) identify several other novel directions in the research that can help understand and improve healthy lifestyle.

Lifestyle-focused research could also focus on a persuasive computing paradigm where technologies could be used to modify human behavior. Personal assistants such as Amazon Alexa, Google Home, etc. are used for providing information and recommendations at the request of a user. These same tools in theory can also be used to provide proactive recommendations for better lifestyles. The feedback loop could be driven by analytics or a recommendation engine. With mobile computing becoming a dominant and pervasive technology, future research could focus on the integration of mobile computing, persuasive computing, and analytics driven decision-making. Analytics has typically been the missing component. A good example of this is work done by Khanal et al. (2014) using a persuasive technology for reducing medical errors and Fukuoka et al. (2011) that demonstrates, through an experiment, how virtual communities support accessibility through mobile phones and how they can assist in controlling type-2 diabetes through lifestyle changes. Such integrative research could help improve medication or program adherence issues. This will also help identify factors that motivate users to stay in the program and barriers that impact sustainability of the program. Given that much of social media exchange is text-based, studies could focus on showing how topic modeling can be applied to mine text to identify dominant themes in a given situation. Aggarwal and Wang (2011) provides a good survey based overview of several approaches that can be applied to mine text data for applications such as social networks.

Several foundations have called for collaborative partnerships between healthcare partners, education, business and other stakeholders to integrate data from disparate sources to improve mental and physical health. With an increasing aging population, there are unique ways in which data about day-to-day caregiving procedures can be captured and relayed to healthcare providers to recommend appropriate strategies for caregivers (Adler 2015). The Center for Disease Control (CDC) and CDC Foundation has now made available city and neighborhood data for over 500 cities in the U.S.A., which has led to some counter intuitive findings (Wojcik 2017) that would have been impossible for public health officials to decipher in the past without access to such data. The social impact of such collaborative efforts brings the necessary and appropriate resources to help communities in much need.

One of the critical challenges for researchers is a lack of access to demographics and health data. RJWF, in collaboration with other partners with similar goals, has organized an

initiative called Health Data for Action (HD4A) with the aim to improve population health through proper lifestyle and a culture of health (Hempstead 2017). While claims data provides a good start point to draw good insights, the addition of data from EHR, socio-economic conditions, lifestyle, wearable fitness tracking, and Internet of Things (IoTs) provide ample research opportunities for general health management such as determining what data is critical for population health management, mechanisms for health risk predictions based on integrating diverse data sources, prescribing meaningful interventions to manage health risks, and building tools to generate customized health alerts. Such efforts can have tremendous influence on the public in terms of making healthy lifestyle choices on a daily basis with or without a healthcare provider's guidance.

### 2.3 Critical Infrastructure

The Department of Homeland Security has identified sixteen different critical infrastructures that need to be secured and made resilient. These are: chemical sector, commercial facilities, communications, critical manufacturing, dams, defense, emergency services, energy sector, financial sector, food & agriculture, government facilities, healthcare, IT sector, nuclear reactor & system, transportation system, and the water & waste water system (Presidential Policy Directive-21 (PPD-21) 2017). Big data and analytics provides several approaches that could be applied to protect these resources through applications such as intrusion detection based on dynamic decision rules, anomaly detection, Supervisory Control and Data Acquisition (SCADA) system protection, and so forth. A good example of such an application is the use of analytics for protecting and improving smart grids.

Smart grids have been playing an increasingly important role in modern energy or electrical systems, creating unprecedented transformation and tremendous opportunities for research, development, and investment (Department of Energy report on the Future of Smart Grid 2015). The World Economic Forum reported in 2015 that an investment of more than \$7.6 trillion would be needed over the next 25 years for the improvement, modernization, and expansion of the global electricity infrastructure (Astarloa et al. 2015). Despite its bright future, the smart grid technology currently faces many challenges, such as accurate prediction of load flow from renewable assets, efficient control of electricity distribution, effective information about and management of power delivery and critical peak loads, and the optimal balance of availability, efficiency, reliability, and costs (Department of Energy report on The Future of Smart Grid 2015). To overcome these challenges, research is needed on how to optimally utilize the smart grids given that the demand from customers varies geospatially at different times of the day/ month/year.

Future studies may be designed to first understand the best conceptual architecture for analysis of data from smart grids

that generate large volume of data from smart meters, SCADA systems, and other accompanying databases. We need to examine the aggregated nature of the data, describing and mapping the energy usage and patterns of usage by census tract and tapestry segment to identify customer groups. Decision support systems can then be built leveraging such analysis that will be useful to decide which geographical regions might yield the greatest decrease in critical peak consumption via dynamic pricing schemes. Such data analytics systems would be useful to individual and large consumers in managing their energy usage and promoting sustainability efforts. Researchers can also use such data to analyze other location-based services such as recovery of utility infrastructure after environmental disasters and storms.

### 2.4 Environment, Energy, and Sustainability

Sustainability is receiving attention among researchers and practitioners alike. While environmental sustainability is linked to phenomenon like climate change, economic sustainability is another aspect of sustainability that is very relevant to organizations that pertains to managing the entire business value chain in an economically sustainable manner. Ryoo and Koo (2013) conduct an empirical study to find that environmental performance is an important predictor of economic performance in organization. Research studies leveraging analytics to address this challenge are only beginning to emerge. For example, Hertel and Wiesent (2013) present an analytics perspective with a decision model for IS investments that promotes energy as well as economical sustainability of a company. This is one of the directions that IS researchers pursuing economic sustainability could adopt.

The U.S. blood system within the healthcare sector can be viewed as an example of a system that demonstrates key aspects of sustainability. Recent disasters caused by hurricane and storm systems such as Harvey, Irma, and Maria have demonstrated that the blood supply chain is not immune to disasters. Various analytics approaches could be applied to study donor prediction, supply chain disruption issues, network of blood flow system, and so forth. RAND Corporation's recent research report quantifies sustainability in a similar context (Mulcahy et al. 2016). A sustainable blood system is one that is able to: (a) maintain or improve desired safety levels for blood and related products, (b) serve a wide range of clinical applications uniformly, and (c) meet blood or related blood product demand in a timely manner without posing undue risk to patient health (Mulcahy et al. 2016). Analytics has a key role to play in maintaining smooth operations while reacting to routine market changes while able to accommodate disruptive changes that might be triggered due to natural disasters, pandemic outbreaks, terrorism related events, and the like. Sustainability of this system can be evidenced in cases of highly disruptive events like 9/11 in

2001, the Sandy Hook shooting in 2012, and Orlando nightclub shooting in 2016. A surge in blood supply demand was observed within a short span of couple of hours at several of these locations. Such adverse events strain local supply system. Network analytics approaches could help evaluate effective approaches to cope under such situations. Studies such as those by Simonetti et al. (2014) and Osorio et al. (2015) show the value of analytics in modeling and managing supply-demand levels in a complex supply chain system like the U.S. blood system.

In other industry segments as well, current challenges and sustainable solutions are being investigated through applications of analytics. For example, in the food safety sector, foodborne illnesses and outbreaks have been shown to have closely aligned characteristics using visual analytics by Ebel et al. (2016). Opportunities for analytics in detecting foodborne diseases and health outbreaks are numerous. Google Flu Trends, one of the early systems for detecting flu outbreaks based search queries by users, while initially highly accurate in 2008, later turned out to miss flu trend in 2013 by as much as 140%, attributed to prevention-aimed search queries, among other reasons (McAfee and Brynjolfsson 2012). This analytics application shows a great example of the potential for analytics applications to leverage user-driven data (e.g., search queries) to conduct predictions about a certain phenomenon (e.g., flu outbreaks) as well as the caveat due to the impact that such applications can have on user behavior (e.g., prevention-focused search queries) which in turn can adversely affect prediction metrics. Regardless, it served as a milestone research in this area. A collaborative study between the New York City Department of Health and Mental Hygiene and Columbia University found that online restaurant reviews from platforms such as Yelp can be used to identify foodborne illness outbreaks linking those to specific restaurants that otherwise would have gone undetected (Harrison et al. 2014). Kaufman et al. (2014) used food sales data from stores to identify contaminated food products using a likelihood-based method. These studies show the potential for using variety of big data sources and employing analytics to solve a key challenge in the food industry pertaining to quick identification of foodborne illness outbreaks, and rollout of appropriate risk mitigation strategies.

From an energy-focused sustainability perspective, smart home initiatives provide a data-rich environment where big data analytics can make an impact for greater societal good. Studies such as those by Hussain et al. (2009) and Palanca et al. (2016) discuss how technologies like wireless sensor network enable an energy conscious smart environment for users through monitoring features and goal-based system design features. Nevertheless, current real-time computational techniques such as Spark Streaming indicate potential for improving these systems several-fold and pave the way for addressing further research challenges in the domain. Future

research studies in sustainability applications may be expected that harness these advanced techniques.

## 2.5 Crowdsourcing and Disaster Relief

Geo-mobile technologies, particularly the use of social media platforms on these technologies, have enabled disaster-stricken people and communities to reach out for help in disaster crises much more easily than ever before (Poblet et al. 2017). Conversely, humanitarian assistance and disaster management agencies and platforms can respond to critical events by applying analytic solutions on crowdsourced data to provide much needed timely help and resources for those in need (Gao et al. 2011). Examples of such crowdsourced data were seen during events such as Hurricane Sandy in 2012 and the Nepal earthquake in 2015.

From a data gathering perspective, analytics solutions rely on both passive and active data sources by harnessing the crowd as reporters of data. Passive data sources include raw data generated by individuals through their location sensors (e.g., GPS receivers) in mobile devices by merely carrying and using these devices. Active data sources, on the other hand, rely on the “power of crowd” to generate data and process information iteratively to produce unstructured data (e.g., tweets, images, videos), and in some cases semi-structured data (e.g., geotags, hashtags). A few disaster response platforms have shown promise through research and practice in the past few years. For example, a system called Ushahidi (Hiltz et al. 2011), deployed internationally in Kenya, Mexico, Afghanistan, and Haiti focuses on generating a real-time crisis map using disparate data sources including crowdsourced data, which is then availed by disaster relief organizations to mobilize aid. GeoCommons (Gao et al. 2011) is a data visualization platform focused on disaster relief coordination and supports temporal analysis. A classification tool developed by researchers from the University of Tokyo was used to analyze real-time tweets to develop a spatio-temporal model for detection of earthquake with very high probability, shown up to 96% (Sakaki et al. 2010).

It is evident that the volume, variety and velocity of the data make the case of disaster relief with crowdsourcing a genuine candidate for applying big data analytics. Equally important is the veracity of the data being reported. People often use social media platforms like Facebook and others during or after disasters to connect with and check the well-being of their near and dear ones. Studies have started emerging on evaluating this critical aspect of data quality and credibility. Gupta et al. (2014) propose a ranking model for scoring tweets using a dataset involving high impact crisis events. Ludwig et al. (2015) propose an application “Social Haystack” that implements a quality assessment approach of crowdsourced data from social media. In the same vein, CrowdHelp (Besaleva and Weaver 2016) is an application that addresses some data

quality issues by providing granular, rich, and user-friendly data reporting mechanisms for victims experiencing medical issues or for bystanders, guiding them to appropriate medical resources, as well as triggering help from relief organization networks as appropriate.

Disaster response management clearly provides a fertile ground to study and implement advanced analytic techniques that can be applied for societal good. Toward that end, researchers should bring to bear crowdsourcing research from other domains that can be extended and applied for disaster relief scenarios. For example, Siering et al. (2016) study how fraudulent behavior can be detected on crowdfunding platforms like Kickstarter by analyzing linguistic and content based cues in unstructured text data about projects.

Crowdsourcing research can inform disaster relief research studies. For example, text mining techniques used in current research can be extended for addressing data quality and credibility issues. Similarly, mechanisms for understanding patterns of diffusion of information in crowdsourced applications (e.g., Waze) using location analytics, can translate to disaster relief scenarios (Pick et al., 2017). In terms of methodologies, previously mentioned new computational AI techniques like Deep Learning also provide novel opportunities to improve prediction performance on big data sources like crowdsourced data. Also, IoT technologies are being leveraged in early-warning systems, as reported in the study by Fang et al. (2015). Another major IoT application, telematics (Karapiperis et al. 2015), which is disrupting the car insurance industry, can be leveraged to enhance disaster relief systems by using location analytics in a different context.

## 2.6 Organizational Resource Management

Organizations constantly work to improve business processes by being efficient and effective in their strategies. Analytics play a huge role in adding value to different parts of any organization's value chain processes (Bedeley et al. 2018). Abbasi et al. (2016) focus on research avenues at the intersection of big data characteristics, the information value chain, and dominant IS research traditions. Any business, profit or not-for-profit, can benefit from good insights based on analytics to formulate differentiation strategies and make informed decisions to add value to their stakeholders. Prior research have proposed models, methods, typologies, domains and factors that lead to successful business analytics applications in organizations (Chen et al. 2012; Evangelopoulos et al. 2012; Wixom et al. 2013; Yeoh and Koronios 2010).

Poverty and social impact analysis (PSIA) an effort of the World Bank, is a “versatile analytical approach to assess the distributional and social impacts of policy reforms on the well-being of different groups of the population, particularly on the poor and most vulnerable” (World Bank 2015). Both quantitative and qualitative methods can be used in PSIA

before or after a policy reform. The outcome hopefully is enhanced policy effectiveness and can contribute to improved national dialogue. In addition, there can be increased accountability and transparency surrounding policies and programs. Stimmel (2014) states there are many nations that are unable to generate sufficient energy efficiently and thus rely on highly polluting fuels to meet basic needs. Collecting and analyzing data related to the type of energy they use coupled with health impacts can assist these nations become more aware of larger problems and implement intervention programs based on energy sustainability to improve overall health. These are good avenues for researchers to collaborate with government and NGOs to assist with appropriate analytics approaches to help identify appropriate reforms and/or interventions that can help developing nations.

Loebbecke and Picot (2015), in their viewpoint paper, examine how digitization and big data analytics (BDA) mechanisms transform business and society. They also outline the potential effects of digitization and BDA on employment, specifically in the context of cognitive tasks. They distinguish five mechanisms by which digitization and BDA complement and replace labor differentially across industry sectors and work processes. Given the pros and cons of each and limited understanding, they call for further research that help tackle the impact of this technology wave.

## 3 Introduction to Special Issue

This special issue presents a collection of seven articles selected after an intensive review process. Authors selected after first round of revision were invited for a one-to-one workshop with the guest editors and were provided additional feedback to further develop the manuscripts.

Plachkinova et al. (2018) proposed a framework at the crossroads of design science and spatial analytics for improving the access to healthcare in the California region as an instantiation of the design artifact. They proposed an improved approach to calculate the healthcare spatial accessibility index based on the floating catchment method. Such a framework could be extended to other application areas societal impact areas that are beyond healthcare for example emergency shelter area accessibility.

Recent research in medical sciences and findings from clinical trials reside in medical literature. Performing an extensive review of literature in a particular medical area is a demanding and involved process. Liu et al. (2018) proposes a semi-supervised approach to make the process of performing scientific medical literature more efficient through the utilization of text analytic approaches.

Popovič et al. (2018), with resource-based view (RBV) as an underlying theory, investigate the impact of big data analytics on three manufacturing cases with varying levels of

analytics including an interpretive qualitative analysis. In addition to anticipated findings of management capability, they also note that empowerment of employees must be integrated in organizations' analytics strategy to yield meaningful outcomes. This is critical for any organization involved in use of analytics to transform operations and production management in manufacturing.

While Popovič et al. (2018) looked into manufacturing sector, Oztekin (2018) develops predictive models using different approaches to understand fund characteristics of organizations. This study investigates various inflows and outflows that can give asset investor in making informed choices about which fund to choose.

Trivedi et al. (2018) present a methodological framework for engagement-based customer segmentation by using signals from social elements to better understand customer profiles. They show how their approach overcomes the limitation of traditional segmentation approaches by using data from an online emotional support service system, which plays a critical role in health management. A different approach is demonstrated by Zhou et al. (2018) study crowdfunding data from Kickstarter in which they use predictors from project descriptions derived using text mining, along with past experience and past expertise as an antecedent for predicting funding success on crowdfunding platforms.

In the paper Schuff et al. (2018), "Enabling Self-Service BI: A Methodology and a Case Study for a Model Management Warehouse", authors Schuff, Corral, St. Louis, and Schymik propose a structured methodology focused on the managing models, particularly at the model formulation stage in predictive analytics projects. The methodology takes the form of a model management warehouse and is instantiated as a dimensional document mart. The dimensions in the warehouse map to various major aspects (e.g., modeling domain, variables, and techniques) of analytical models being built and facilitate querying the model base for model selection and related tasks using this information. The feasibility and the efficacy of the approach has been demonstrated using a modeling process situated in the US Division of Fiscal and Actuarial Services. The selected case involves 53 geographically separated and disparate units where this methodology has been demonstrated to work effectively. It is evident that the methodology has broad applicability and scalability that can positively impact model management efforts in societal applications at large. This study demonstrates the need for making the analytics and modeling process easier for various applications so that wider applicability could be seen in the future.

## 4 Conclusion

Innovative solutions utilizing big data and analytics approaches often require connectedness among various data

sets. This also brings important issues of ethics and privacy. Regulations and policy compliance are to be given due while requiring data integration, data misuse and misinterpretation while addressing societal problems that have grand impact.

We discussed various big data and analytics approaches that can be applied to study issues that have major societal impact. We have identified some key aspects that are inspired from the papers included in this special issue. Examples of specific societal applications include healthcare, lifestyle, critical infrastructure, environment, energy and sustainability, disaster response, crowdsourcing and resource management. Significant future analytics driven research needs to happen in each of these areas for visible societal impact.

## References

- Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17(2), i – xxxii.
- Adler, (2015). Bringing the technology revolution to caregiving. RWJF Report. [http://www.rwjf.org/en/culture-of-health/2015/09/bringing\\_the\\_technol.html](http://www.rwjf.org/en/culture-of-health/2015/09/bringing_the_technol.html), Retrieved March 2017.
- Aggarwal, C. C., & Wang, H., (2011). Text mining in social networks social network data analytics (pp. 353–378): Springer.
- Andreu-Perez, J., C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong and G. Z. Yang, "Big data for health," in *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1193–1208, July 2015.
- Astarloa, B., Critchlow, J., & Miller, M. (2015). The future of electricity attracting investment to build tomorrow's electricity sector. *World economic. Forum, 2015*.
- Baesens, B., Bapna, R., Marsden, J. R., & Vanthienen, D. (2016). Transformational issues of big data and analytics in networked business. *MIS Quarterly*, 40(4), 807–818.
- Bedeley, R., Ghoshal, T., Iyer, L. S., & Bhadury, J. (2018). Business analytics and organizational value chains: a relational mapping. *Journal of Computer Information Systems*, 58(2), 151–161. <https://doi.org/10.1080/08874417.2016.1220238>.
- Besaleva, L. I., & Weaver, A. C. (2016). Applications of social networks and crowdsourcing for disaster management improvement. *Computer*, 49(5), 47–53. <https://doi.org/10.1109/MC.2016.133>.
- Bresnick, J. (2016). "Smart Big Data is Key to Population Health, Value-Based Care", *Healthcare Analytics*, [healthitanalytics.com/news/smart-big-data-is-key-to-population-health-value-based-care](http://healthitanalytics.com/news/smart-big-data-is-key-to-population-health-value-based-care), Retrieved: August 20, 2017.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
- Department of Energy (2015) *The Future of Grid: Evolving to meet America's need* (report prepared by New West Technologies).
- Ebel, E. D., Williams, M. S., Cole, D., Travis, C. C., Klontz, K. C., Golden, N. J., & Hoekstra, R. M. (2016). Comparing characteristics of sporadic and outbreak-associated foodborne illnesses, United States, 2004–2011. *Emerging Infectious Diseases*, 22(7), 1193–1200. <https://doi.org/10.3201/eid2207.150833>.
- Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, 21, 70–86.
- Fang, S., Xu, L., Zhu, Y., Liu, Y., Liu, Z., Pei, H., & Zhang, H. (2015). An integrated information system for snowmelt flood early-warning

- based on internet of things. *Information Systems Frontiers*, 17(2), 321–335. <https://doi.org/10.1007/s10796-013-9466-1>.
- Fitzgerald, M. (2015). When health care gets a healthy dose of data. *MIT Sloan Management Review*.
- Fukuoka, Y., Kamitani, E., Bonnet, K., & Lindgren, T. (2011). Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: A focus group analysis. *Journal of Medical Internet Research*, 13(3), e49. <https://doi.org/10.2196/jmir>.
- Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, 26(3), 10–14. <https://doi.org/10.1109/MIS.2011.52>.
- Gruetzemacher, R., Gupta, A., (2016) Using Deep Learning for Pulmonary Nodule Detection & Diagnosis, Twenty-second Americas Conference on Information Systems (AMCIS 2016), San Diego.
- Gupta, A., & Sharda, R. (2013). Improving the science of healthcare delivery and informatics using modeling approaches. *Decision Support Systems*, 55(2), 423–427.
- Gupta, A., Li, H., & Sharda, R. (2013). Should I send this message? Understanding the impact of interruptions, social hierarchy and perceived task complexity on user performance and perceived workload. *Decision Support Systems*, 55(1), 135–145.
- Gupta, A., Kumaraguru, P., Castillo, C., & Meier, P. (2014). TweetCred: Real-time credibility assessment of content on twitter. In L. M. Aiello & D. McFarland (Eds.), *Social informatics. SocInfo 2014. Lecture notes in computer science, vol 8851* (pp. 228–243). Cham: Springer. [https://doi.org/10.1007/978-3-319-13734-6\\_16](https://doi.org/10.1007/978-3-319-13734-6_16).
- Harrison, C., Jorder, M., Stern, H., Stavinsky, F., Reddy, V., Hanson, H., & Balter, S. (2014). Using online reviews by restaurant patrons to identify unreported cases of foodborne illness — New York City, 2012–2013. *Morbidity and Mortality Weekly Report (MMWR)*, 63(20), 441–445.
- Hempstead, (2017). “New effort will give researchers access to valuable health datasets”. RWJF Report. <http://www.rwjf.org/en/culture-of-health/2017/04/effort-will-give-researchers-access-to-health-datasets.html>, Retrieved August 20, 2017.
- Hertel, M., & Wiesent, J. (2013). Investments in information systems: A contribution towards sustainability. *Information Systems Frontiers*, 15(5), 815–829. <https://doi.org/10.1007/s10796-013-9417-x>.
- Hiltz, S. R., Diaz, P., & Mark, G. (2011). Introduction. *ACM Transactions on Computer-Human Interaction*, 18(4), 1–6. doi:<https://doi.org/10.1145/2063231.2063232> <https://energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid> Last Visited, Sept 5th, 2017 [https://www.smartgrid.gov/files/Read\\_Ahead\\_Document\\_-\\_Central\\_Region\\_Workshop.pdf](https://www.smartgrid.gov/files/Read_Ahead_Document_-_Central_Region_Workshop.pdf), Last visited: Sept 5<sup>th</sup> 2017.
- Hussain, S., Erdogen, S. Z., & Park, J. H. (2009). Monitoring user activities in smart home environments. *Information Systems Frontiers*, 11(5), 539–549. <https://doi.org/10.1007/s10796-008-9124-1>.
- Kalgotra, P., Sharda, R., & Paiva, W. D. (2017). Analysis of a sequence of events in healthcare. *Journal of the Midwest Association for Information Systems*, 2017(2), 9–23.
- Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*, 18(2), 233–235. <https://doi.org/10.1007/s10796-016-9641-2>.
- Karapiperis, D., Birnbaum, B., Brandenburg, A., Castagna, S., Greenberg, A., Harbage, R., & Obersteadt, A. (2015). Usage-Based Insurance and Vehicle Telematics: Insurance Market and Regulatory Implications. *National Association of Insurance Commissioners & The Center for Insurance Policy and Research*, (March), 1–80. Retrieved from [http://www.naic.org/documents/cipr\\_study\\_150324\\_usage\\_based\\_insurance\\_and\\_vehicle\\_telematics\\_study\\_series.pdf](http://www.naic.org/documents/cipr_study_150324_usage_based_insurance_and_vehicle_telematics_study_series.pdf)
- Kaufman, J., Lessler, J., Harry, A., Edlund, S., Hu, K., Douglas, J., et al. (2014). A likelihood-based approach to identifying contaminated food products using sales data: Performance and challenges. *PLoS Computational Biology*, 10(7), 1–10. <https://doi.org/10.1371/journal.pcbi.1003692>.
- Khanal, P., Vankipuram, A., Ashby, A., Vankipuram, M., Gupta, A., Drumm-Gurnee, D., Josey, K., Tinker, L., & Smith, M. (2014). Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. *Journal of Biomedical Informatics*, 51, 49–59.
- Liu, J., Timsina, P., & El-Gayar, O. (2018). A comparative analysis of semi-supervised learning: The case of article selection for medical systematic reviews. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9724-0>.
- Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems*, 24(3), 149–157. <https://doi.org/10.1016/j.jsis.2015.08.002>.
- Ludwig, T., Reuter, C., & Pipek, V. (2015). Social haystack: Dynamic quality assessment of citizen-generated content during emergencies. *ACM Transactions on Computer-Human Interaction*, 22, (17), 1–17,27. <https://doi.org/10.1145/2749461>.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60–6, 68, 128. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23074865>
- Mulcahy, Andrew, A., Kapinos, K., Briscoe, B., & Uscher-Pines, L. (2016). Toward a Sustainable Blood Supply in the United States: *An Analysis of the Current System and Alternatives for the Future*. Retrieved from [http://www.rand.org/pubs/research\\_reports/RR1575.html](http://www.rand.org/pubs/research_reports/RR1575.html)
- Osorio, A. F., Brailsford, S. C., & Smith, H. K. (2015). A structured review of quantitative models in the blood supply chain: A taxonomic framework for decision-making. *International Journal of Production Research*, 53(24), 7191–7212. <https://doi.org/10.1080/00207543.2015.1005766>.
- Oztek, A. (2018). Information fusion-based meta-classification predictive modeling for ETF performance. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9704-4>.
- Palanca, J., Del Val, E., Garcia-Fornes, A., Billhardt, H., Corchado, J. M., & Julián, V. (2016). Designing a goal-oriented smart-home environment. *Information Systems Frontiers*, 1–18. <https://doi.org/10.1007/s10796-016-9670-x>.
- Pick, J. B., Turetken, O., Deokar, A. V., & Sarkar, A. (2017). Location analytics and decision support: Reflections on recent advancements, a research framework, and the path ahead. *Decision Support Systems*, 99, 1–8. <https://doi.org/10.1016/j.dss.2017.05.016>.
- Pickering, B. W., Dong, Y., Ahmed, A., Giri, J., Kilickaya, O., Gupta, A., Gajic, O., & Herasevich, V. (2015). The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: A pilot step-wedge cluster randomized trial. *International Journal of Medical Informatics*, 84(5), 299–307.
- Plachkinova, M., Vo, A., Bhaskar, R., & Hilton, B. (2018). A conceptual framework for quality healthcare accessibility: A scalable approach for big data technologies. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9726-y>.
- Poblet, M., Garcia-Cuesta, E., & Casanova, P. (2017). Crowdsourcing roles, methods and tools for data-intensive disaster management. *Information Systems Frontiers*, 1–17. <https://doi.org/10.1007/s10796-017-9734-6>.
- Popović, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms’ high value business performance. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9720-4>.
- Presidential Policy Directive-21 (PPD-21). (2017). Critical infrastructure security & resilience. <https://www.dhs.gov/critical-infrastructure-sectors>. Accessed 5 Sept 2017.



- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4–21. Robert Wood Johnson Foundation; 2014. Health Information Technology in the United States: Progress and Challenges Ahead.
- Ryoo, S. Y., & Koo, C. (2013). Green practices-IS alignment and environmental performance: The mediating effects of coordination. *Information Systems Frontiers*, 15(5), 799–814. <https://doi.org/10.1007/s10796-013-9422-0>.
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. Proceedings of the 19th international conference on world wide web, 851–860. <https://doi.org/10.1145/1772690.1772777>.
- Schuff, D., Corral, K., St. Louis, R. D., & Schymik, G. (2018). Enabling self-service BI: A methodology and a case study for a model management warehouse. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9722-2>.
- Siering, M., Koch, J.-A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: the role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems*, 33(2), 421–455. <https://doi.org/10.1080/07421222.2016.1205930>.
- Simonetti, A., Forshee, R. A., Anderson, S. A., & Walderhaug, M. (2014). A stock-and-flow simulation model of the US blood supply. *Transfusion*, 54(3pt2), 828–838. <https://doi.org/10.1111/trf.12392>.
- Stimmel, C. L. (2014). *Big data analytics strategies for the smart grid*, 1st edn. Boston: Auerbach Publications.
- Trivedi, N., Asamoah, D. A., & Doran, D. (2018). Keep the conversations going: engagement-based customer segmentation on online social service platforms. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9719-x>.
- Wilkerson, G. B., Gupta, A., Allen, J. R., Keith, C. M., & Colston, M. A. (September 2016). Utilization of practice session average inertial load to quantify college football injury risk. *Journal of Strength & Conditioning Research*, 30(9), 2369–2374.
- Wilkerson, G. B., Gupta, A., & Colston, M. A. (2018). Mitigating sports injury risks using internet of things and analytics approaches. *Risk Analysis*. <https://doi.org/10.1111/risa.12984>.
- Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive*, 12(2), 111–123.
- Wojcik, O. (2017). “The 500 cities project: New data for better health”. RBWJ Report. [http://www.rwjf.org/en/culture-of-health/2017/02/the\\_500\\_cities\\_proj.html](http://www.rwjf.org/en/culture-of-health/2017/02/the_500_cities_proj.html), Retrieved August 20, 2017.
- World Bank (2015). Poverty and Social Impact Analysis (PSIA). <http://www.worldbank.org/en/topic/poverty/brief/poverty-and-social-impact-analysis-psia>
- Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of Computer Information Systems*, 50(3), 23–32.
- Zhou, M., Lu, B., Weiguo, F., & Wang, G. A. (2018). Project description and crowdfunding success: an exploratory study. *Information Systems Frontiers*, 20(2). <https://doi.org/10.1007/s10796-016-9723-1>.

**Ashish Gupta** is a Harbert College Advisory Council Faculty Fellow and an Associate Professor of Analytics in Raymond J. Harbert College of Business at the Auburn University. He has been a Visiting Research Scientist at the Mayo Clinic Rochester, Visiting Associate Professor in Biomedical Informatics at the Arizona State University and research affiliate with University of Tennessee Health Science Center in Memphis. His research interests are in the areas of data analytics, healthcare informatics, sports analytics, social media, organizational and individual performance. His recent articles have appeared in journals such as *MIT Sloan Management Review*, *Journal of Biomedical Informatics*, *IEEE Transactions of Health & Biomedical Informatics*, *Information Systems*

*Journal*, *European Journal of Information Systems*, *Decision Support Systems*, *Information Systems Frontiers*, and *Communications of the Association for Information Systems*. His research has been funded by several agencies and private enterprises. He has published 5 edited books. He has a PhD in MSIS from Spears School of Business at Oklahoma State University.

**Amit V. Deokar** is an Assistant Professor of Management Information Systems in the Manning School of Business at the University of Massachusetts Lowell. Dr. Deokar received his PhD in Management Information Systems from the University of Arizona. He also earned a MS in Industrial Engineering from the University of Arizona and a BE in Mechanical Engineering from VJTI, University of Mumbai. His research interests include data analytics, enterprise data management, business intelligence, business process management, and collaboration processes. His work has been published in journals such as *Journal of Management Information Systems*, *Decision Support Systems (DSS)*, *The DATA BASE for Advances in Information Systems*, *Information Systems Frontiers*, *Business Process Management Journal (BPMJ)* and *IEEE Transactions*. He is currently a member of the editorial board of DSS and BPMJ journals. Dr. Deokar was recognized with the 2014 IBM Faculty Award for his research and teaching in the areas of analytics and big data. He currently serves as the Chair of the AIS Special Interest Group on Decision Support and Analytics (SIGDSA).

**Lakshmi S. Iyer** is Professor of Information Systems and Director of the Applied Data Analytics Graduate Programs at the Walker College of Business, Appalachian State University. Her research interests are in the area of emerging technologies & its impact on organizations and users, and social inclusion in computing. Her research work has been published in or forthcoming in *Communications of the AIS*, *Journal of Association for Information Systems*, *European Journal of Information Systems*, *Communications of the ACM*, *Decision Support Systems*, *eService Journal*, *Journal of Electronic Commerce Research*, *Journal of Computer Information Systems*, *Information Systems Management*, *Journal of Global Information Technology and Management*, and others. She is a Board member of Teradata University Network, past-chair of the Special Interest Group in Decision Support and Analytics (SIGDSA, formerly SIGDSS). She has served as a Guest Editor for *Communications of the ACM*, the *Journal of Electronic Commerce Research*, and *Information Systems Frontier*. She is also co-editor of *Annals of Information Systems Special Issues* (2015 and 2017) on “Reshaping Society through Analytics, Collaboration, and Decision Support: Role of BI and Social Media” and “Analytics and Data Science: Advances in Research and Pedagogy”.

**Ramesh Sharda** is the Vice Dean for Research and Graduate Programs, Watson/ConocoPhillips Chair and a Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. He has coauthored two textbooks (*Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, 4th Edition, Pearson and *Business Intelligence and Analytics: Systems for Decision Support*, 10th edition, Pearson). His research has been published in major journals in management science and information systems including *Management Science*, *Operations Research*, *Information Systems Research*, *Decision Support Systems*, *Decision Sciences Journal*, *EJIS*, *JMIS*, *Interfaces*, *INFORMS Journal on Computing*, *ACM Data Base* and many others. He is a member of the editorial boards of journals such as the *Decision Support Systems*, *Decision Sciences*, and *Information Systems Frontiers*. He also serves as the Faculty Director of Teradata University Network. He received the 2013 INFORMS Computing Society HG Lifetime Service Award, and was inducted into Oklahoma Higher Education Hall of Fame in 2016.

**Dave Schrader** worked for 24 years as a marketing and advanced development software engineering director at Teradata. He was involved in the Teradata® Unified Data Architecture™ Big Data and Active Intelligence™ initiatives. He led the marketing of Teradata's Active Intelligence real-time data warehousing initiatives. Throughout his career, Schrader assisted Teradata customers to learn how to use both traditional and big data to create analytical insights and predictive models. He continues to help Teradata as a board member of the Teradata University Network, giving talks at universities and helping them with curriculum

development for data analytics programs. He continues as a board member of the Teradata University Network, giving talks at universities. He currently focuses on helping faculty and students get connected with their own athletic departments to do sports analytics projects. He gave more than 50 talks at 25 schools to more than 1700 students in 2017. Schrader holds a Ph.D. in computer science from Purdue University, has published in the areas of customer management and business intelligence, and is a popular conference speaker.