Big Data, What Is It, Its Limits and Implications in Contemporary Life

Andreas Karaoulanis

MScEng, Aristotle University, Greece; MBA, Blekinge Institute of Technology; Masters in Decision Support and Risk Analysis, Stockholm University; Masters in Information Systems, Linnaeus University, Sweden

Abstract

This paper summarizes the arguments and counterarguments within the scientific discussion on the issue of big data in the contemporary world in terms of what big data is, how it functions, how it can be leveraged towards the common good and what limitations can prevent it from transforming its power to a springboard for development and growth.

The main purpose of the research is to get a first grip on what big data is and its implications positive or negative to our world. Under this prism, the author of this paper tries to encapsulate in this literary analysis some of the many ways that big data can help us in many industries like the health and care, insurance etc. and to underline the importance of its handling due to severe confidential data breaches, like in the case of the U.S. last elections.

Systematization of literary sources and approaches for solving the problem of big data's limitations indicate that big data need to be handled with extreme care and caution. As sensitive personal information is involved, the companies which use big data in order to understand their customers' approach and way of thinking towards them in order to increase their sales funnel, need to handle it in a very cautious way, especially when they outsourcing that procedure to third party companies.

This paper presents the results of an in-depth literary analysis on the subject, which showed that big data is undeniably an important part of our societies and that it has specific characteristics (i.e. speed, volume etc.) which make its analysis a quite challenging procedure, which needs to involve new techniques like data mining. The results of the research can be useful for the researcher of the future in terms of examining the connection between big data, artificial intelligence and personal information. This subject is critical and needs to be addressed in a coherent way as the advent of artificial intelligence and machine learning will arise new issues on how intelligent machines will handle personal information in the years to come.

Keywords: Big Data, limits, implications, contemporary, use, analysis.

JEL Classification: C55, C8.

Introduction

People nowadays are living in a complex world getting signals from so many directions all around them. Their way of living, what they do, what they like, where they travel, their social habits, their professional attitudes and in brief their whole lives are consisting of a huge amount of data.

About 7 billion people on this planet live their own lives on a multidimensional level as they are parents, husbands, friends, employers or entrepreneurs etc. and all their activities, thoughts, moments captured in a camera, reactions and everyday stories can be depicted online in several ways.

All the information that we can find about people's lives and not only as we can find information about almost everything on this planet, creates a humongous amount of data, the biggest of all times.

This kind of huge amounts of data spreading so rapidly needed a specific technology to support it. As Infarate (2014) indicated, Google in 2004, was the first corporation which deployed the appropriate technology needed in order to process the huge amounts of data that we coming to its servers on a daily basis after the billions of queries that were made online to its bigger search engine in the world.

The way we are using big data is crucial as it can be handled for the good or for the bad. Issues like privacy are emerging i.e. every time we face big data and how we save it or leverage its benefits. Big corporations

need to address issues like safety in order to have the proper approach to the issue and to keep their customers safe. Also, accuracy and validity issues need to be addressed by the analyzing companies in order to prevent false results which can be misleading.

On the other hand, big data can be a powerful tool, which could alter the face of the world via its analysis, which can deliver outstanding results and solve complicated issues that weren't able to be solved the previous years, like in disease forecasting or in pattern analyzing etc.

It is obvious from the about that the big data topic is a very important and interesting and in need of further investigation of the researcher of the future, not to mention that nowadays it has to play a very important role in so many daily activities and concerns.

The author of this paper strives to deliver a brief description of what big data is, its usage in modern societies and its limitations. In that way, we can have a first impression and understanding of this huge issue, which has so many implications in our contemporary world.

What Big Data Is?

It is more than obvious from the above that data nowadays is really big and is getting bigger as we speak.

Some of the most important and to the point definitions of Big Data are the following:

According to Pense (2014), Big Data is defined not only by the amount of the information that delivers but also by its complexity and by the speed that it is analyzed and delivered.

Wu et al. (2014) indicated that Big Data has to do with large- volume, complex, growing data sets which comes from multiple and autonomous sources.

Iafrate (2015), defined Big Data as the data which follows the four "V"s, which are volume, variety, velocity, value.

As Clarke (2016) underlines, the main characteristics of Big Data are volume, velocity and variety (Laney, 2001, cited in Clarke, 2016: 77). Clarke (2016) continues that such statement leads us the definition that Big Data is the data that is too big, too fast or too hard to be processed by the contemporary tools (Jacobs, 2009: 44, cited in Clarke, 2016: 77).

As we can see there are a lot of definitions of Big Data, but in my opinion, the following is the most coherent and the one which can clearly describe the real omni-face of what we call "Big Data".

So, according to Gordon (2013: 12), Big Data can be defined as a multi-faced data which combines the following characteristics:

  I. *VOLUME:* Data, in order to be "Big" need to present a sufficiently large amount, so as to require special considerations from the people who are handling it (Gordon, 2013).

 II. *VARIETY:* The data in question need to have a subsequent variety of forms. Under this prism, Gordon (2013), underlines that Big Data need to be of multiple types coming from multiple sources. So, structured data (coming from tables or objects where the metadata is well defined), semi-structured data (i.e. documents where the metadata is contained internally),unstructured data (photos, videos or any other form of binary data) can be the most important types that we encounter in the Big Data concept (Gordon, 2013, p.12).

III. *VELOCITY:* Dats, in order to be "Big" need to be produced at high rates in juxtaposition with "stale" data which is not valuable (Gordon, 2013: 12).

IV. *VALUE:* Data needs to create quantifiable value to the organization, which uses it (Gordon, 2013).

 V. *VERACITY:* The data need not only to be correct, but its correctness need to be able to be assessed (Gordon, 2013: 12).

So, from the above analytical definition of Big Data, we can easily understand that in order to distinguish the difference between data and Big Data, we need to see if the data in question is coming in huge amounts, from different sources under different forms/ types, it is coming quickly, not to mention rapidly, if it is creating value for the users and if we can assess its correctness (Gordon, 2013).

Big data is something different than the data we used to know. Is data that exists at a humongous volume, which due to its size is unmanageable by the ordinary computer systems. It is a data, which arrives at a unique

velocity, especially nowadays, a velocity which doesn't permit any kind of ordinary analysis and investigation (Brennan et al, 2015). Automatically, we can understand that there is a gap which is created in terms of how to access and handle big data and this is where data science methods come to take place by managing big data and gaining insights from it, which will help us understand potential patterns and create value to our everyday activities etc. (Brennan et al., 2015).

What Big Data Can Do for Us?

After reviewing what big data is in order to understand its core nature, which essentially is described by the five Vs as we saw in the previous chapter, we need to see how can we leverage its potential and create value for us. Companies nowadays try to leverage big data, but that is not always easy as it increases exponentially. And this is only one part of the problem. Big data, in its majority, increases not only in volume but also in forms that are difficult to handle. Such forms can be information like i.e. call center voice data, social media content and video entertainment and clickstream data from the web (Barth et al., 2012).

So, organizations have to use the modern sophisticated software in order to be able to handle big data. The really important thing is that when they are able to handle it that gives them a competitive advantage against other organizations which don't.

As they use big data, they differentiate their approach from companies that still use the traditional data analysis, because they do not have to look at the past in order to find patterns of data, which will help them to predict future behaviors. Instead, the consider data in terms of flows and processes, something which helps them to take decisions and act quickly (Barth et al., 2012).

As they try to leverage big data, they tend to hire big data scientists for its analysis, something which places them one step ahead from the companies which use traditional data analysis (Barth et al., 2012).

Finally, the result of such approach is that they can move their IT results into their core business and operations in order to produce new products and to implement new processes, things that bring to their company added value (Barth et al., 2012).

Big data is not what we used to see as a pool of statistics. On the contrary, we can see it like an ongoing process, which when it comes to organizations and customers, it has to do with the information which involves the customers' choices (Davenport, 2014). That is why we need a more continuous approach in terms of sampling, analyzing and acting on big data (Davenport, 2014).

There are many examples of how big data is used in customer-centric organizations, i.e. financial institutes tend to create a detailed map of the so-called "customer journey" through their websites, call centers, tellers and other branch personnel in order to be able to understand which path their customers take through the organization and how this path affects their purchasing behavior (Davenport, 2014).

Big data is very useful to companies which use an omnichannel approach of their customers' in order to create a clear enough picture of the way their customers move when they interact with their organizations. So, under that prism, such companies try to pinpoint their customers' actions via i.e. company's social media page, the company's site, the customers' visits to the brick and mortar company's facilities, their interaction with a potential call center etc. In that way, by analyzing this big data, they will be able to offer to their customers a 360 degrees customer satisfaction experience, something which will help them a lot in terms of creating loyal customers.

Of course, as everything in life, big data can be used for the good or for the bad. The use of big data as a way of systematic monitoring of people or groups of people, by means of personal data systems in order to regulate or govern their behavior, is called "dataveillance" and it refers to the "dark" side of big data in terms of how people use it (Degli Esposti, 2014).

Big data can be used in so many ways. It is literary, everywhere. Organizations are exploring how large volume data can usefully be deployed in order to create value for people, businesses, communities and governments (McKinsey Global Institute, 2011, cited in George et al., 2014). Only to think that we can use it analyzing genetic sequences, count social media interactions, create health records, phone logs, governments' records, introducing new forms of journalism (cf visualization), helping in crime risk assessments etc. (Randall, 2018).

Another way big data can influence today's life is via healthcare. As big data "by definition" gives us the ability to handle massive data volume very quickly the right infrastructure, analytic tools, visualization approaches, workflows etc. will help the healthcare industry to manage big data in the proper way, something

which will add to it value via the improved outcomes and the lower costs (Roski et al., 2014). This can happen by eliminating adverse events and reducing readmissions to hospitals (Schaeffer et al, 2017).

According to a study that was implemented in the US health care industry (Schaeffer et al. 2017), although the results indicated that health care has so much to gain from the big data implementation, it also has to face the fact that some new expenses will occur like the hiring of the personnel needed to handle this new form of data in an adequate way (Scaeffer et al., 2017).

Of course issues like the patients' confidentiality, need to be addressed and confronted by the healthcare people (Roski et al., 2014).

Farming and agriculture is another field where big data can help us very much. As we all know, farmers all around the globe are always dependent to weather. Data was here to help them overcome this paramount obstacle, but now the advent of big data and its analysis via data science will help farmers to grab weather predictions in a more accurate and swift way (McSharry et al., 2015).

Among many, one more industry that will be benefited by the advent of big data is the insurance industry. Actuaries need to calculate risks and costs of certain insurance policies and because the exponential growth of the industry the last years the leverage of big data came to help them so much (McSharry et al., 2015).

Big data can help actuaries in calculating cash flows, demographics of their insured people and to target their groups in accordance with their product that needs to be promoted, calculating the probability of changing events in their customers' life, their professional risks etc. (McSharry et al., 2015). They calculated them already but the last years the data gathered went big and that made almost inevitable the handling of this big data in a proper way.

Another known example of the use of big data is the Cambridge Analytica[1] (Randall, 2018), which claimed 5,000 data points per person in terms of their demographics like lifestyle, gender, age, location. Consumer behavior, political affiliations etc. It was also included the psychological profiling of these persons in order to assess their motivation and decision-making. Of course, we all heard the rumors that it was involved in Trump's election in the USA, so there was a controversial issue there about its usage (Randall, 2018).

Another industry that has been vastly influenced by the advent of big data is the education one. Big data has proven to be very important in terms of reframing coursework and grading development as well as in data-driven classrooms (usually virtual ones). As we can now understand, what students learn and to what weight, we can address the specific lessons needed to specific students' categories respectively to their needs (Intellipaat, 2018).

The entertainment industry is one that everybody knows and use on an almost daily basis. Platforms like Spotify, Amazon Prime etc. are some of the specific examples of data used in the entertainment world. For example, Spotify which as we know is an on-demand music providing platform, gathers data from all of the users from around the world and after its analysis, gives customized recommendations to its customers in order to reach their needs and interests (Intellipaat, 2018).

Finally, another industry that benefits the most out of the big data analysis is the banking one. Issues like the misuse of credit or debit cards, the archival of inspection tracks, the venture credit hazard treatment, business clarity, customers' statistics alteration, IT action analysis, money laundering etc. are only some of the so many that big data analysis helps the banking sector to deal with its contemporary everyday problems (Intellipaat, 2018).

It is obvious from the above, that big data can be used in so many industries and daily life activities in order to create value, but we also need to know that there are certain limitations in this use. We are going to deal with these limitations in the following chapter.

Big Data Limitations

We understood from what we saw in the previous chapter that big data can be used in many directions in the entire spectrum range of all social activities and can generate value for so many industries. Big data usage is critical in terms of how accurate it is. According to Channel Futures website (2017), the accuracy of big data

---

[1] Cambridge Analytica: "Cambridge Analytica is a company that offers services to businesses and political parties who want to "change audience behavior". It claims to be able to analyze huge amounts of consumer data and combine that with behavioral science to identify people who organizations can target with marketing material. It collects data from a wide range of sources, including social media platforms such as Facebook, and its own polling" (The Guardian, 2018).

has to do with mainly two paragons, firstly it depends on how you use it and secondly on what you are using it for. Permit me to add that it depends on how you analyze it too, in terms of which methods you are using, the sample etc.

According to Croft (2014), there are three major questions that needed to be answered in order to understand the potential limits of big data nowadays. These questions are the following:

1. How big data is collected, stored, analyzed and reported? An essential question, especially for policymakers and technologists who strive to create and inform policy.
2. Is big data capable of changing the world? Its role will be vital in terms of helping policymakers and i.e. doctors in predicting and saving lives or we are going to consider it just another metric?
3. Will big data serve to behavior modification based on the knowledge of collection?
   (Croft, 2014).

Certain limitations exist in big data analytics, which is used as subcomponents for descriptive, predictive and prescriptive tasks, often trained in using machine learning (Varshney, 2016). So, one can argue that when such analytic components will become parts of large sociotechnical systems, it is often difficult for humans to predict how such systems will react (Varshney, 2016). This is a limitation that we need to bear in mind as several implications might occur for which we need to be prepared.

A similar problem is that while big data is driving the use of the algorithm in governing mundane but mission-critical tasks, algorithms become increasingly autonomous and some can say invisible, while it becomes more and more difficult to detect and scrutinize their impartiality status. But algorithms might fail in their implication as they can systematically cultivate an i.e. discrimination pattern in several domains like in history, politicals, several practices etc. (Janssen et al., 2016). Such bias might create severe problems in societies. On the other hand, of course, this is not a problem that big data and/ or algorithms create as behind algorithms is the human who created them in order to analyze data and result in the patterns needed to describe situations or even predict the future.

Pitrelli (2017), stresses that Stuart Allan and Joanna Redden addressed issues related to privacy, web disinformation, the opacity of big data algorithms, the quality of big data, which meets the epistemological needs of citizen science. Such issues are important in terms i.e. science journalists who are required to create new information gathering and analyzing strategies in order to be able to contextualize and to broaden the debate on the risks and the opportunities that arise from the enormous data sets that citizens produce (Pitrelli, 2017).

Financial risk analysis is another sector where big data can have vast influence in the next years. Brammertz et al. (2014) underline that big data has value in that sector too as long as it is structured, something, which is quite critical for this case. As they stress, structured big data needed in terms of representing financial contracts as algorithms that produce state contingent cash flows. As the specific industry lacks such standards, big data is here to stay and its limits in terms of its implementation in that sector are yet to be known (Brammertz et al., 2014).

According to Dezzain.com website (2017), the main three big limitations of big data are:

I. **Misinterpretation of big data:** While big data can be a powerful tool in unveiling the users' actions, it cannot help you in terms of discovering why the users in question thought or acted in that way. This fact per se creates a knowledge gap, which can be critical and lead to misinterpretation of their motives. This is crucial to both marketers and sellers who are trying to identify their customers' needs in order to reach them.

II. **Security limitations:** As companies continue to collect data, they have the big responsibility to safeguarding this data. Recent laws like GDPR are a good example of how seriously such obligations need to be taken. Another vivid example is when companies use third-party groups or agencies to have their data analyzed. Data leak from them or directly from the company can lead to lawsuits and big fines and in some cases even with creating extensive damage to the company's brand name.

III. **The outlier effect:** In data mining process of data analyzing we face in almost every analysis we make the so-called "outliers". "An *outlier* is an observation that lies an abnormal distance from other values in a random sample from a population" (Itl. nist.cov, n.d.). In simple words, data that is away from our sample can be critical and be taken out of consideration, at least at the level needed, just because it is in the outlier area. This is an incapability of the analyses systems that we use nowadays. A well-known example of such case is the Google flu trends, a program that was created by the Google engineers in order to provide a real-time monitoring of global flu cases by gathering data from organic search with

terms that involved the word flu in order to help researchers track flu cases on a global scale. The whole program ended to a disaster due to Google's algorithms inability to properly predict search behaviors. That was and still is, the company's most spectacular failure to date (Dezzain.com, 2017).

Another field where big data faces some limitations is the so-called "sentiment classification" which is the task of assigning a positive or negative label to a text segment (Filippas et al., 2017). This classification is a key component of mainstream applications like reputation monitoring, sentiment summarization and item recommendation (Filippas et al., 2017). Filippas et al. (2017) stress that after a certain point, the predictive performance of this method tends to almost stop and additional data has little or no benefit at all.

Finally, among the author's concerns for the following years is "how A.I.[2] can leverage the use of big data in terms of interfering with people's privacy". This is a very complicated issue, which needs to be addressed in a separate research paper, as it is not in the scope of this one.

Conclusions

In this paper, we discussed what big data is, its limits and implications positive or negative in contemporary societies. Issues like the misinterpretation of big data, security issues, the outlier effect and the sentiment classification are some of the most important limitations that big data faces nowadays. On the other hand, big data can be extremely useful in contemporary life in fields like health, trade, insurance etc.

Big data is here to stay. Information, thus data is spreading in exponential ways that we cannot even imagine. Although we can find so many ways to use that data flow in order to improve our lives, limitations and possible dangers can also be met.

New tools like data mining need to be improved and be implemented in order to be able to harvest the benefits of big data. On the antipodes, data breaches need to be taken seriously in order to provide a safe environment for the people of this world. The data manipulation is a big issue and laws like GDPR need to go even deeper in order to protect privacy.

How can we use big data in terms of creating value without interfering in a negatively way in people's lives is a big issue that we need to address in a future research. Big data, as almost everything in life, can be used in a good and in a bad way. The advent of A.I. in the years to come can create many questions in terms of how A.I. can manipulate big data, but this is something we need to discuss in depth in another paper, as it cannot be spread in only a few lines.

References

1. Barth, P., Bean, R. & Davenport, T. H. (2012). *How 'Big Data' Is Different,* MIT Sloan Management Review.
2. Brammertz, W. & Mendelowitz, A. (2014). Limits and Opportunities of Big Data For Macro-Prudential Modeling of Financial Systemic Risk, ACM, pp. 1.
3. Brennan, P.F. & Bakken, S. (2015). Nursing Needs Big Data and Big Data Needs Nursing. *Journal of Nursing Scholarship, 47*(5), 477-484.
4. Channel futures (2017). Recognizing the limitations of big data and data analytics. [online] Available at: https://www.channelfutures.com/channel-futures/recognizing-limitations-big-data-and-data-analytics. Accessed 14 October 2018.
5. Clarke, R. (2016). Big data, big risks. *Information Systems Journal, 26*(1), 77-90.
6. Croft, C. (2014). The Limits of Big Data. *SAIS Review of International Affairs, 34*(1), 117-120.
7. Davenport, T. H. (2014). How strategists use "big data" to support internal business decisions, discovery and production. *Strategy & Leadership, 42*(4), 45.
8. Degli Esposti, S. (2014). When big data meets dataveillance: the hidden side of analytics. *Surveillance & Society, 12*(2), 209-225.
9. Dezzain.com (2017). The three big limitations of big data. Available at: http://www.dezzain.com/business/the-three-big-limitations-of-big-data/. Accessed 14 October 2018.
10. Filippas, A. & Lappas, T. (2017). Strength in Numbers: Using Big Data to Simplify Sentiment Classification. *Big data, 5*(3), 256-271.
11. George, G., Haas, M. & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal, 57*(2), 321-326.

---

[2] A.I.: Artificial Intelligence

12. Gordon, K. (2013). What is Big Data? *ITNOW, 55*(3), 12-13.

13. Intellipaat (2018). 7 Big Data Examples – Application of Big Data in Real Life [blog]. Available at: https://intellipaat.com/blog/7-big-data-examples-application-of-big-data-in-real-life/. Accessed 06 November 2018.

14. Itl.nist.gov, n.d. What are outliers in the data? [online]. Available at: https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm. Accessed 14 October 2018.

15. Iafrate, F. (2015). "What is Big Data" in John Wiley & Sons, Inc, Hoboken, NJ, USA, pp. 1-11.

16. Janssen, M. & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly, 33*(3), 371-377.

17. McSharry, P. & Thomas, R. (2015). *Big Data Revolution: What farmers, doctors and insurance agents teach us about discovering big data patterns,* John Wiley & Sons Inc, GB.

18. Pence, H. E. (2014). What is Big Data and Why is it Important? *Journal of Educational Technology Systems, 43*(2), 159-171.

19. Pitrelli, N. (2017). Big data and digital methods in science communication research: opportunities, challenges and limits. *Journal of Science Communication, 16*(2), 1-4.

20. Randall, D. (2018). Lecture 6: Big Data, 18HT – 5IK502 – contemporary issues in IS research and development, Linnaeus University Department of Informatics, Unpublished.

21. Roski, J., Bo-Linn, G.W. & Andrews, T.A. (2014). Creating Value In Health Care Through Big Data: Opportunities And Policy Implications. *Health Affairs, 33*(7), 1115.

22. The Guardian (2018). *What is Cambridge Analytica? The firm at the centre of Facebook's data breach.* [Online] Available at: https://www.theguardian.com/news/2018/mar/18/what-is-cambridge-analytica-firm-at-centre-of-facebook-data-breach. Accessed at 10 October 2018.

23. Schaeffer, C., Booton, L., Halleck, J., Studeny, J. & Coustasse, A. (2017). Big Data Management in US Hospitals: Benefits and Barriers. *The health care manager, 36*(1), 87-95.

24. Varshney, L. R. (2016). Fundamental Limits of Data Analytics in Sociotechnical Systems. *Frontiers in ICT, 3*, 1-7.

25. Wu, X., Zhu, X., Wu, G. & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*(1), 97-107.