Journal of Technology Management for Growing Economies

Journal homepage: https://tmg.chitkara.edu.in/

Big Data and Its Applications

Nishu Bali[1], Deepika Chaudhary[2] and Anshu Singla[3]

[1,2,3]*Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab-140401, India*

[1]nishu.bali@chitkara.edu.in
[2]deepika.chaudhary@chitkara.edu.in
[3]anshu.singla@chitkara.edu.in (Corresponding Author)

ARTICLE INFORMATION

ABSTRACT

In times when everything is online, one thing which is common in every application is the use of data. Data is being generated every second, when applications are generating exponentially larger data sets every second; it's the big data which comes into effect. The major objective of this paper is to state the meaning of big data, figure out various ways of making use of the enormous data efficiently. Further this paper will also focus on the applications of Big Data in multiple segments: Finance, Banking and Securities and Health Care Sector.

1. Introduction

Big data refers to the huge amount of data generated from different sources. Everything is online these days therefore the technological and societal progress have made it difficult to maintain, query and manage this ever increasing complex and heterogeneous data. People now days are surrounded by various devices for collaborating, socializing and career building as well. Applications like Facebook, Twitter, YouTube and LinkedIn produce a huge amount of data. The amount of data generated belongs to different research areas and fields and the amount is so huge that it is almost impossible to deal with it by existing conventional tools. Data generating giants like Google, WhatsApp, Facebook, Twitter and other social networking sites are continuously creating trillions of data in a day. Google has to its credit 1.2 trillion searches daily and more than 40000 search queries per second. Similarly, WhatsApp users exchange up to 65 billion messages in a day. Twitter users exchange over half a million tweets every minute (Petrov, 2020). In 2012 a research was conducted by the American Multinational Corporation (AMC) which has estimated the growth of data from the year 2005 till 2020, they concluded that growth rate of data will be by a factor of 300 and therefore will be doubled every two years. Similarly IBM also estimated that 90% of the emerging data has been generated in the past two years (Marr, 2015). In addition, Oracle also estimated the similar pattern .This exponential growth of data had led to the "Big Data" phenomenon. The next section focuses on the meaning and definitions of Big Data.

1.1. Meaning and Definition of Big Data

Big data is defined as complex and large data sets being generated from multiple heterogeneous data sources including sensors, social media, IoT and administrative services (Petrov, 2020). In all these scenarios it is not only about the amount of data generated that is important but also how well the data is being used and analyzed by the organization for better decision making and for strategic business moves. All big businesses feel the need to manage unstructured data which is a need of the hour. Big Data as per the literature reviews can be divided into four dimensions: velocity, volume, value, and variety (Petrov, 2020). Volume is all about the size of data; from tera to petabytes, velocity is about the frequency of occurrence of data, variety is about the type of data, structured, semi-structured and unstructured data and the major challenge is of maintaining the value of data such that which data to keep and which to discard. This rich data is of utmost important for the researchers as this data, if stored and analyzed properly, can lead to multiple benefits and provide valuable insights. For example with

the data generated from social media sentiment analysis and opinion mining can be a research area. This research can also help in market research segment where valuable insights can be provided to customers. In the section below we present the various application areas of big data.

1.2. Areas for Big Data Application

Big data has found its applications in almost every sphere of life. As data is something which is relevant in all the researches of today, every field is going to be dependent on wise and timely analysis and evaluation of data. The coming section of the paper is going to highlight some of the areas in which big data has shown significant contribution.

1.2.1. Banking and Securities

Banking industry has an important role to play in the financial health of a nation. In the present scenario of internet and its advancements, the use of online apps and portals by clients is on an increase for various banking services. Thus, banks of today have to take care of not only offline dealing with the client demands but also for the security of online applications or possible frauds. Thus the challenges that a banking industry has to face include securities fraud, card fraud detection, archival of audit trails, enterprise credit risk reporting, trade visibility, customer data transformation, social analytics for trading, IT operations analytics, and IT policy compliance analytics, among others (Hassani, Huang & Silva, 2018). Big data can play a vital role in managing these sectors. The condition of financial market is being scrutinized by Securities Exchange Commission (SEC) using Big Data Network analytics and natural language processors are being utilized by the latter to control illegal trading activities in financial market. Other types of analysis like risk analysis, anti-money laundering, demand enterprise risk management, fraud mitigation are also heavily dependent on Big Data. The continuous developments in the field of big data will influence the efficiency of data mining applications in banking sector (Hassani, Saporta, Silva, 2014).

1.2.2. Health Care Sector

Health care sector generates a huge amount of structured and unstructured data everyday owing to the growing health problems in people today. Patients, nowadays, are not only dependent on doctors for their treatment but are also referring to various mobile applications for their various healthcare needs. These mobile apps are designed to keep track of the symptoms of various diseases, their cure, preventive measures and other health related information. The data generated by these devices and through other health care mediums and hospitals remains unutilized in

an efficient form in reducing the cost of healthcare and in providing better healthcare benefits across the globe. Big data can be of great help in this sector in helping the doctors and other health workers in identifying the patterns from the huge dataset of patients and recommend better ways of dealing with the upcoming challenges in healthcare sector. Medical internet of things (MIOT) is being utilized for integrating the data from various devices with telemedicine and telehealth for reducing amount of time required to bring together all pieces of data and do required analytics on them (Dimitrov, 2016).
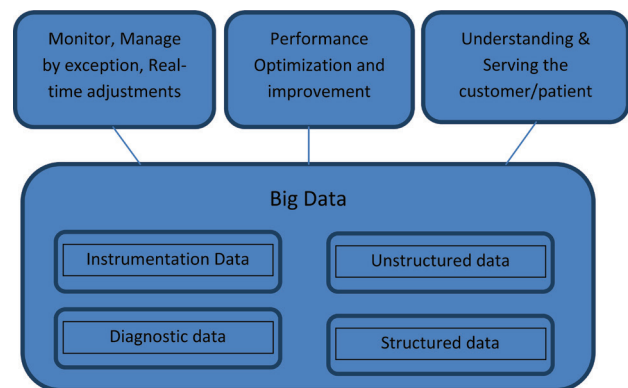

Figure 1: Big Data in Health Sector.

1.2.3. Role of Big Data in COVID-19

In the present scenario of COVID-19, Big data played a vital role. Huge data of patients suffering from this deadly virus was very difficult to be analyzed quickly and efficiently. Big data helped in identifying patterns in the patients that helped in COVID-19 timely detection and recovery. Also, It was Big data startup: BluDot, based in Toronto, that perceived the unusual pneumonia cases in Wuhan, China in December 2019 (Bragazzi et al., 2020). Here also, Big data algorithm was used to pull out data from various sources like airline ticketing data, health records and other such sources to predict the spread of this disease. Thus Big data can help the governments in facing and preventing more such challenges in future (Brüssow, 2020).

1.2.4. Education

COVID-19 has brought a huge paradigm shift in the way education is provided to the students. Education through Online classes has become the only way. To efficiently analyze the level of education the students are given, their learning experiences, to measure the impact of this type of education on the mind and stress level of learner, a thorough analysis is required which is only possible through Big data analytics. As the data generated through online classes has to be efficiently analyzed for assessing the learner's scores, Big

data tools can be of great help in the area (Sin & Muthu, 2015).

1.2.5. Big Data for Smart Cities

Advancements in the field of information technology are increasingly getting utilized for the betterment of human life. One such concept is the concept of smart cities. Smart city is a concept of integrating technologies to improve living style of citizens of the city in different aspects such as power consumption and production, transportation, building smart solutions to day to day problems of the people and at the same time maintaining the sustainability of resources for the future generations (Allam & Dhunny, 2019). Smart cities generate huge amounts of data which is utilized by Big data to enhance the applications of smart cities. Also using Big data analytics, expansion policies regarding smart city services, resources and areas can be formulated. Big data systems can also be utilized for storing, processing and mining the smart city applications efficiently to produce information which can be used to enhance different smart city services. Artificial intelligence in combination with Big data can provide solutions to challenges like pollution and congestion and at the same time can ensure decent livelihood by providing clean, healthy and conducive environment (Al Nuaimi, Al Neyadi, Mohamed & Al-Jaroodi, 2015; Zanella, Bui, Castellani, Vangelista & Zorzi, 2014).
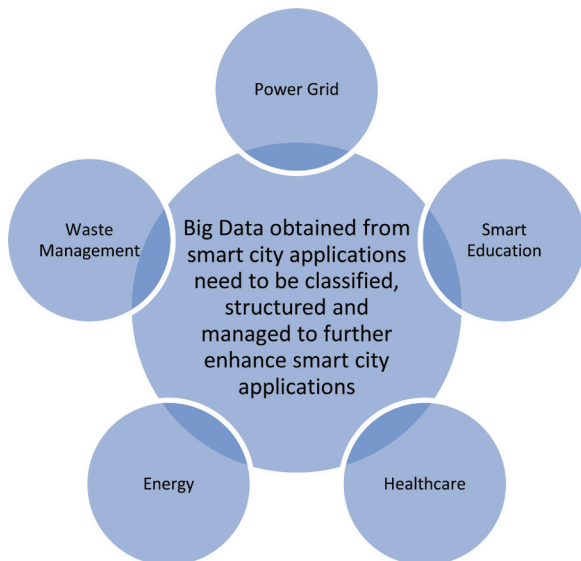

Figure 2: Smart City in relation with Big data.

1.2.6. Smart Farming

The advancements in the field of agriculture has led to tremendous growth in the volumes of data generated. The use of smart machines and sensors on fields has made data pertaining to crops readily available for further analysis.

Smart Farming is a technique in which the information and communication technology join hands for a better farm management (Zanella, Bui, Castellani, Vangelista & Zorzi, 2014). There is a significant role of Big data in smart farming which include better predictive modeling for crops, risk analysis in case of crop failure, sensor data analytics. An appreciable change in scope and organization of farms can be brought by the inclusion of Big data in the field. Access to large amount of information and decision making capabilities of Big data can bring a radical change in the farm management and farmers can be better decision makers while assessing their gains or predicting their loss. Advancements in field of machine learning can be combined with big data for more gains in the field (Wolfert, Ge, Verdouw & Bogaardt, 2017).

1.2.7. Biomedical Text Mining and Named Entity Recognition (NER)

In Biomedical domain people are trying to integrate mathematics along with computer science and various biological tools in order to perceive the knowledge from the data sets. This can be stated as a process of data analysis for deriving high quality information which is previously unknown and very useful for masses. The current research in this area can be divided into two sub sections which is extracting and mining. NER is very important and critical to Big data mining which includes classification of DNA, RNA and cell_line and cell_types (Barmpounakis et al., 2015; Song, Jo, Park, Kim & Kim, 2018). Timely and accurate mining on enormous medical literature can identify novel techniques to better understand and diagnose human diseases and can lead to better treatments (Ye, Tafti, He, Wang, & He, 2016).

In the next section we present the broad challenges faced in this direction and a few steps on how to overcome those challenges.

2. Challenges

The major challenges regarding the maintenance of Big Data can be classified into four categories.

If we talk about all the key challenges with respect to the domain mentioned in the paper; **Volume** can refer to storing of data used to manage real time traffics, real time capturing of vital signs for patients, storage of unstructured data related to hyper spectrum images and their analysis. To process and manipulate these data sets the required algorithms should be provided with the mechanism of scalability. **Velocity** refers to the continuous generation of data at a very rapid rate and the difficulty lies in real time streaming and processing of this data. **Variety** deals with the heterogeneity of data which

means data coming from various sources and in various formats along with also to record locations from which these data is coming is challenging. One more challenge which has been added to the list above is **Veracity** which is related to the privacy part. This is required to hide user identity and to cater this requirement leads to a robust and secure algorithm design. Further to these challenges the captured data is also noisy and obfuscated or malicious which again has to be handled very maturely (SAS Insights, 2020).

Figure 3: Challenges in Big Data management.

Conclusion

In this paper, applications of Big data in various fields has been discussed. The power of Big data tools in providing good analysis of huge amount of data has shown its application in almost every field of today. The advancements in the field of information and technology are continuously generating volumes of data which, if fully utilized, can lead to more fruitful inventions and discoveries and can make the life of humans more and more comfortable. The challenges of Big data in the form of 4 Vs (volume, velocity, variety and veracity) need to be tackled in a manner so that the efficiency of Big data can be explored fully for the betterment of human kind.

Authorship Contribution

Nishu Bali: Collected data and compiled it in the form of manuscript, revised the manuscript according to the suggestions from other authors.

Anshu Singla: Conceived the idea, guided for the creation of manuscript, contributed to relevant data, reviewed the drafts and provided valuable suggestions.

Deepika Chaudhary: Contributed to the idea and data collection for the manuscript, provided valuable suggestions and contributed to the formation of manuscript.

Funding

Conflict of Interest

There is no conflict of interests of authors with anyone.

References

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 25. https://doi.org/10.1186/s13174-015-0041-5

Allam, Z., & Dhunny, Z.A. (2019). On big data, artificial intelligence and smart cities. *Cities*, 89, 80-91. https://doi.org/10.1016/j.cities.2019.01.032

Barmpounakis, S., Kaloxylos, A., Groumas, A., Katsikas, L., Sarris, V., Dimtsa, K., Fournier, F., Antoniou, E., Alonistioti, N., & Wolfert, J. (2015). Management and control applications in Agriculture domain via a Future Internet Business-to-Business platform. *Information Processing in Agriculture*, 2(1), 51–63. https://doi.org/10.1016/j.inpa.2015.04.002

Bragazzi, N.L., Dai, H., Damiani, G., Behzadifar, M., Martini, M., & Wu, J. (2020). How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*, 17(9), 3176. https://doi.org/10.3390/ijerph17093176

Brüssow, H. (2020). The Novel Coronavirus - A Snapshot of Current Knowledge. *Microbial Biotechnology*, 13(3), 607–612. https://doi.org/10.1111/1751-7915.13557

Dimitrov, D.V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, 22(3), 156-163. https://doi.org/10.4258/hir.2016.22.3.156

Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. *Big Data and Cognitive Computing*, 2(3), 18. https://doi.org/10.3390/bdcc2030018

Hassani, H., Saporta, G., & Silva, E.S. (2014). Data Mining and Official Statistics: The past, the present and the future. *Big Data*, 2(1), 34–43. https://doi.org/10.1089/big.2013.0038

SAS Insights (2020). History of Big Data. Retrieved from https://www.sas.com/en_in/insights/big-data/what-is-big-data.html

Marr, B. (2015). Big data predictions and reflections. Retrieved from https://www.ibmbigdatahub.com/blog/big-data-predictions-and-reflections

Petrov, C. (2020). 25+ Impressive Big Data Statistics for 2020. Retrieved from https://techjury.net/blog/big-data-statistics/#gref

Sin, K., & Muthu, L. (2015). Application of Big Data in Education Data Mining And Learning Analytics - A Literature Review. *ICTACT journal on soft computing,* 5(4), 1035-1049. https://doi.org/10.21917/ijsc.2015.0145

Song, H. J., Jo, B. C., Park, C. Y., Kim, J. D., & Kim, Y. S. (2018). Comparison of named entity recognition methodologies in biomedical documents. *Biomedical engineering online*, 17(Suppl 2), 21-34. https://doi.org/10.1186/s12938-018-0573-6

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.J. (2017). Big data in smart farming–a review. *Agricultural Systems*, 153, 69-80. https://doi.org/10.1016/j.agsy.2017.01.023

Ye, Z., Tafti, A.P., He, K.Y., Wang, K., & He, M.M. (2016). Sparktext: Biomedical text mining on big data framework. *PloS one*, 11(9), e0162721. https://doi.org/10.1371/journal.pone.0162721

Zanella, A., Bui, N., Castellani, A., Vangelista, L. & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1). 22-32. https://doi.org/10.1109/JIOT.2014.2306328

Journal of Technology Management for Growing Economies

Chitkara University, Saraswati Kendra, SCO 160-161, Sector 9-C, Chandigarh, 160009, India