Improving Healthcare Using Big Data Analytics

*R.Reena, **Shalini. S, ***K.Rajalakshmi
Associate Professor, Department of Computer Science & Engineering
*Prince Shri Venkateshwara Padmavathy Engineering College, Ponmar Chennai-600127.
**Prince Dr.K.Vasudevan College of Engineering & Technology, Chennai.
***Faculty, Prince Shri Venkateshwara Arts and Science College, Gowriwakkam.

Abstract:. Our daily lives in today's world are rapidly advancing never quenching ones thirst. The fields of science, engineering and technology are producing data at an exponential rate leading to Exabyte(s) of data every day. Big data helps us to explore and re-invent many areas not limited to education, health and law. The primary purpose of this paper is to provide an in-depth analysis in the area of Healthcare using the big data and analytics. The main purpose is to emphasize on the usage of the big data which is being stored all the time helping to look back in the history, but this is the time to emphasize on the analyzing to improve the medication and services. Although, many big data implementations happen to be in-house development, this proposed implementation aims to propose a broader extent using Hadoop, which just happen to be the tip of the iceberg. The focus of this paper is not limited to the improvement and analysis of the data; it also focuses on the strengths and drawbacks compared to the conventional techniques available.

Index Terms: Behavior Analysis, Big data, Clinical Analysis, Data mining, Descriptive Analytics, Healthcare, Location based analytics, Predictive, Prescriptive, Patient centric, Structured & Unstructured data.

1 INTRODUCTION

Data is a powerful resource which is found in many forms. Big data do not have a universal definition while it is discussed in different ways. The term Big data is referred to describe the exponential growth of the data flow in various sectors which is too large to process using the available traditional database and software techniques. Often big data is presumed to be scary, yet it is an explosion in the field of information. It helps to perform various analytics, which can make an impact on the economic growth, creating opportunities, improving efficiency over other organizations. This significant heap of data often defined as three-dimensional [1] namely Volume, Velocity and Variety where some even define with Veracity.

Volume: Data volume is a contribution by various factors. It can be transactional data, which is being used through the years, or the data flow over the social media. The volume of the data is the total quantities of the mass data within an organization. The volume of data generated in an organization increases daily at an unpredictable rate, which can be in peta bytes and zeta bytes on the production activities and the type of the organization.

Velocity: This refers to the data in the total data transmitted currently in an organization or in motion. The speed of the data that an organization produce process and analyzes normally keep on accelerating. It influences the creation and delivery of the data from one point to the next. It is often time-sensitive.

Variety: The variety, which is diverse in forms, type of data and its origin. It defines the complexity of the data, and the Occurrences of data. It is in any form like structured, semi-structured and unstructured data. Some forms of structured data are the Numerical data, traditional databases, business information and unstructured data like Audio, Video and Pictures.

Veracity: Veracity, which is composed of the data that the organization is uncertain. It analyzes levels of forms of data credited on reliability. Organizations enactment of strategies to ensure quality and reliable data is normally hindered by factors such as weather and customer's reactions and purchasing decisions.

2 WHAT IS BIG DATA ANALYTICS

-Information is the oil of the 21st century, and analytics is the combustion engine.‖ – Peter Sondergaard, Gartner Research. A significant growth is distinguished when it can make a positive impact. The data collected in various repositories by various organizations, the data produced by the individuals can make a difference only if we can analyze and utilize the data properly. In other words, without proper analytics, data will be just a resource but not a utilized resource. In addition, here the term big data does not only talk about the volume of data but also the power of the data. The data sets are large and complex, challenging the current techniques to analyze and capture the outcomes. In order to overcome and make decisions in the rapidly growing business sector, the big data analytics sifts through the data to uncover the hidden pattern, realize unknown correlations, understanding the market trends, customer preferences and other useful business information.

Descriptive analytics helps to illustrate the picture about the previous history using the business intelligence and data mining [0]. As we all know, experience teaches a lot. Using this analytics helps to provide a way to draw an approach to accomplish the goal.

The **predictive analytics** using the huge datasets helps to improve the customer experience increasing the outcomes comparative to the conventional business strategies. It helps to analyze large volumes of transactional data, unstructured data at once providing the outcomes helping to predict the future. Predicting the future, based on the available datasets has been a tough task all through the computing days until date. Business Intelligence programs of this kind help to compute the data streams at a larger extent including the social media content, shopping experiences, daily user activities and survey reports.

Prescriptive analytics can be termed as a suggestion tool

The system provides advices based on the outcomes it has generated, by learning the patterns, previous approach and the inputs from the data (structured and unstructured). This mechanism also allows analyzing and prescribing based on the outcomes of any other research on the same activity by inputting the data from other research, it helps to connect the dots and provide solutions based on both the attributed works. Since, we cannot rely on machine completely, which is a human invention; it is only considered as the possibilities. However, based on the outputs, an extensive research can be performed considering the other possible solutions towards a problem.

3 BIG DATA IN HEALTHCARE

Defining the volume of the data, the type of data, and the entity limitations are very wide. Combining the complete health care data is very large, which refers as -big data‖ but not as huge, as we think that the organization cannot handle its data. Most of the healthcare providers have not faced any harder situation to handle the data for them; however, it is always good to foresee the technology improvements and implementation, which can help them. According to the McKinsey Global Institute, better targeting of preventative healthcare messages to the right population at the right time could save $70-100 billion [0]. As a result, Hadoop data processing is the one of the best choice to go with at the current trends. The computational capabilities of Hadoop processing will be able to stimulate the mathematical methods available currently, medical research approaches to increase the outcome quality. Although big data might not be their case, but the discovery process to find new techniques to analyze the data they have, increase the accuracy of the experimental results, provide various mechanisms to find the quality of data is always at the highest priority. As a result, Hadoop data processing is the one of the best choice to go with the current trends. The computational capabilities of Hadoop processing will be able to stimulate the mathematical methods available currently, and medical research approaches to increase the outcome quality. -Most of the data systems are for billing, and they aren‘t used to improve the quality of care,‖ explains Jason Jones, executive director for clinical intelligence and decision support at Kaiser Permanente, a health care provider and not-for-profit health plan that serves approximately 9.1 million members in 8 states and the District of Columbia [0]. The emerging generic health care systems usually save and manage EMR (Electronic Medical Record), PHR (Personal Healthcare Record) [0], Laboratory Information System (LIS), biomedical data, biometrics data, and genomic data [0] which can be the invaluable sources to generate the outcomes. These various data sources help to process and analyze the data with various characteristics. Processing such massive data sets using the Hadoop technology will helps us not only to process quickly compared to the traditional database solutions which are being used currently, but also provides an extra edge to analyze the data characteristically.

4 OBJECTIVES

The objective of my project is to propose a feasible computing solution using the big data and analytics. It aims to foster the research, availability and accessibility in the field of health care. This project also provides measurable benefits providing the ground realities to improve the field of health care. It targets to benefit the society with advanced computation techniques to analyze and provide patient-centric health care. The proposed objectives with detailed description are as below.

Clinical Decision Support

The Clinical Decision Support (CDS) aims to increase the quality of health care services enhancing the outcomes. The primary focus of the system is to provide the right information to the right people, proper customized health care management process but not limited to clinical guidelines, documentations, and diagnosis. This system enables doctors, pharma, patients and other individuals to know the information in any specific medical related information.

Disease Management

This system enables to analyze various diseases, its evolution using the laboratory tests. The analytics help to improve the accuracy to find out the information enabling to improve the outcomes such that medical solutions can be determined. This requires the data support from various organizations, medical repositories and the individuals.

Patient Matching

With the help of prescriptive analytics using the big data solutions, patient-centric medical approach is developed. It aims to explore the previous disease management systems, the approach to cure the patient, symptoms while fighting the disease etc. By exploring and analysis, the outcomes could be clear enough to treat a patient based on the symptoms instead of a generic disease-based management.

Lifestyle Analytics

The proposed system helps to provide the health care solutions based on various methods but also life style of the individuals. In order to prevent the medical accidents and increase the accuracy towards the disease detection, lifestyle of the individual plays an important role. This system provides the picture in a wide range of various lifestyle activities done by the individuals. This helps to study the impacts and the causes.

5 DATA MATCHING

In order to implement the system without any errors, as accurate as possible and achieve greater results; it is required to deal with multiple repositories, and medical organizations.

6 DATA SECURITY

Throughout the process, Privacy will be the primary concern. Since, medical records contain various personal information; data access is limited to the authorized users. Besides, data is stored in a highly secured server including regulations such as Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Data Protection Act and encryption techniques.

7 DATA PYRAMID

The pyramid depicts the architecture and explains about the Data Management. It clearly gives the picture of the data flow starting from the raw data along with its types through the Hadoop eco-system and the analytic engines to achieve the final goal of the system.

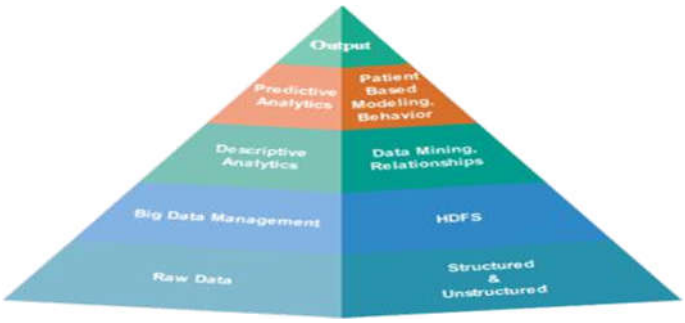


Figure 1: *Data Pyramid*

8 DATA LIFECYCLE

The lifecycle of a system defines the process that has ability to impact on the efficiency providing problem solutions.

Figure 2: *Data Lifecycle*

Data Collection

The very first stage of the process is collecting the data from various repositories, sources and storing it in Hadoop Distributed File System (HDFS). Data can be both structured and unstructured like clinical analysis, pharmaceutical purchases, patient history, reports, medical emergencies, health index, and social media content, wearable devices and so on. The impact on the quality can be measured with the collection process.

Data Pre-processing

As we collect the data from different sources and store it into a common storage, this stage helps to cleanse the data, sort accordingly for further analysis. All the missing values, columns or data will be discarded. Preparing the data plays an important role in moving the process ahead as raw data cannot be processed and might not help to achieve any results. In this process, any junk data will be discarded.

Data Reduction & Transformation

Thus far we have been staging & preparing the incoming data.This stage helps to process the data from the pipeline by reducing least prominent data (or) columns. Data transformation aims to transform the data applying various mathematical & compression algorithms. Transforming the data will not lose any data rather helps to sort out to process further easily. Various analytics and analysis functions can be implemented on the transformed data.

Data Analytics

Data Analytics plays the major role in analyzing the data by drawing conclusions. It interprets the data for research, science and business decisions. Using these outcomes, business models can be improved, unknown correlations can be made.

Data Output

The statistical, analytical information is presented based on the inputs given. The generated reports are interpreted appropriately in various formats for the user to view, monitor or print. The output data is the meaningful information that guides and helps to achieve the goal. In order to store the output data for any further use, data storage is required. Similarly, it is important to maintain security, integrity and data access control specified throughout the lifecycle.

9 USE CASES

Healthcare is one of the most important industries which seek improvement providing the planet Earth with better solutions to Live life happily; the use of big data and analytics is widely spread across every field we can think of. As we discussed about this emerging technology, efficiency, and its uses, it is clear to understand its significance in the healthcare industry. Healthcare, nowadays has become very complex that, the computation required in understanding the DNA is complicated. In the whole economic system, healthcare is a part of it. The services of healthcare don't necessarily limit to the medication but also in the prevention, rehabilitation, palliative care at various regional levels, which might be domestic or global. Increase in the wearable devices for fitness enables the individuals to access complete electronic reports of their health providing a global access through a secure protocol with valid credentials. In addition, wearable fitness devices can help to analyze the patterns of the patient throughout their day increasing the accuracy towards their health care. The doctors will be given a chance to view the changes in their metabolism and body changes. Putting these values in a graph would enable to understand an in-detail pattern during various times of the day. It is no wonder that in the future, hospitals might become important for the intensive care, and there can be expected a decline in the appointments by the patients for general checkup or follow up check-ups. Many researches have already stated that many people are showing interest towards their health even more from the day of the evolutionary wearable fitness devices. Although, the data produced from these devices are structured, the analytics can compare the values in terms of days, months and years. Predicting the future health, limited to a person-specific improves to fight against the disease and a better care. Research and testing in areas like protein, genomic and DNA is very large in scale. To understand the protein variants and their effects, laboratory experiments are essential which are not available for an exact reason so far until date. To distinguish between the protein variants containing 1000 amino acids which would have 19,000 variants, and test cases for such variants carried out by changing by just one amino acid at a time [0].

hug
mar
big
app
area

Healthcare expenditure (% Total of GDP)		Healthcare GDP per capita	
2012	2013	2012	2013
17.05	17.1	8845.18	9145.83
5.41	5.57	321.69	366.86
10.28	10.3	4787.1	3965.58
11.27	11.3	4716.59	5006.5
9.27	9.12	3594.71	3597.92

data generated every day. Implementation of big data analytics will not only change the outcomes and the future perspective towards the improvements, it would help the medical researchers/analysts to move towards an Evidence- Based Medicine. Providing evidence based medicine increases the accuracy towards the treatment of a patient enabling quick recovery, which will lead to overcome the generic medications that are been traditional until date. With the focus of treatment based on the events and the evidences, the rate of clinical accidents can be reduced.

.

Figure 3: *Online Healthcare Monitoring System for improved Analytics.* [000000]

10 Economy, Expenditures and Statistics of Healthcare

The economic spending of the countries around the world is quite different from each other yet only to be an increase. There also exists a demand to increase the economic allocations and spending. The growth in the GDP expenditures can tell us how much a country spends on various aspects to improve. The statistical data take from the World Bank is below for better understanding.

Table 1: *Total GDP for 2012 & 2013 with GDP growth*

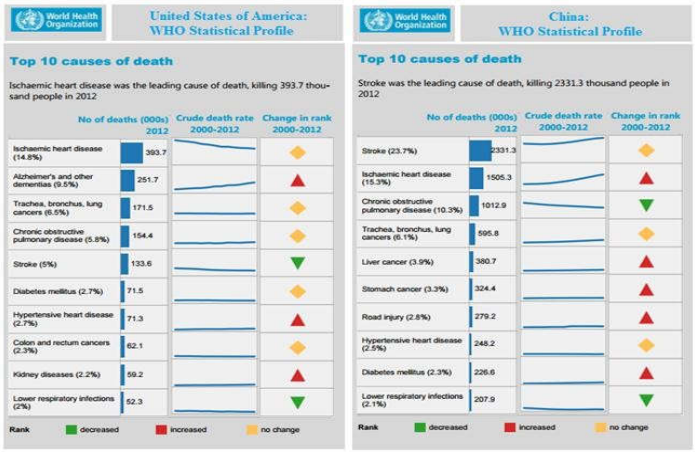
Currency Type: USD	Total GDP (in billion)		GDP Growth (annual %)
Country	2012	2013	
United States	16163	16768	2.22
China	8462	9491	7.68
Japan	5954	4920	1.61
Germany	3533	3730	0.11
United Kingdom	2615	2678	1.66

In table 1, GDP of top five countries are considered for a quick overview on the spending's and their annual growth. It is clear that the United States tops the chart and United Kingdom stands at fifth position, but the annual growth of the country was not directly proportional to the total GDP. This is a clear example, how various factors affected its annual growth irrespective of their allocations.

Table 2: *Percentage of healthcare expenditure and per capita expenditure for total GDP*

Table 2 gives a picture of the healthcare expenditures by the same 5 countries in the years 2012 and 2013. The increase from the previous year is at a low rate, however when the spending on an individual is compared to be better in real time compared to the ground reality. These figures talk on an average spending, if the spending is real for every single individual, it is still at a very high cost. Although, they are among the high economic countries, the statistical data gives a picture on the decline of the importance and growth towards healthcare. It is not said, the availability is less but there are very high chances of low accessibility by a common person. The services provided by the health care professionals who have become costly to approach are one of the reasons[0]. As the advantage is leveraged using the analytics, it is possible for every country to take a leap in this field. If we consider the same statistical data for any country that is not forward in their economy, it is obvious to expect a steep decline towards the care of health. Such places need to be uplifted for the betterment of the society. In this process, there should be a continuous and rapid research where the data production will be beyond the limit that a human can think. As the volume of the data increases, the complexity to analyze increases exponentially which can also take years of time to achieve the results with every combination possible.

Figure 4: *WHO Statistical Profile - Top 10 causes of death in USA and China for the year 2012[0].*



The above picture published by the World Health Organization (WHO) gives an overview of the most common causes of death in the countries USA and China in the year 2012. Of all, the overall rank in change is disappointing. It is again clear that the accessibility is very low. If we consider 'Stroke', it is not that the person was not able to reach the hospital but it is very likely to doubt if the person has proper knowledge to deal with such situations. Accessibility indirectly refers to the healthcare education. Therefore, the big data and analytics are not only about the research, scientific findings and understanding the patterns or predict the future, but also a solution for the basic problems like health education, awareness and quick accessibility.

11 Conclusion

One of the essential budget oriented talk is healthcare and its cost, mainly in the countries having very less economic status, areas with bad health hygiene, increasing births and aged people. In this study, we propose to analyze the healthcare using the big data analytics specifically to any given geographic location and the data available. In addition, moving towards the big data storage and solutions would provide an efficient solution in contrast to the traditional storage solutions. Any further research can easily extend the system to improve the facilities and services.

References

[1] IBM, (2015, September 25). Big data at the speed of business: Big data and analytics, Retrieved from URL http://www-01.ibm.com/software/data/bigdata/

[2] IBM. (2013). Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics, Retrieved from document. TIW14162USEN, IBM Corp.

[3] Sa-kwang Song et.al. (2013). Prescriptive Analytics System for Improving Research Power, IEEE 16th International Conference on Computational Science and Engineering, pp.1144-1145.

[4] Good read: Scientists Save Healthcare (But They‘re Not Fro(n.d.). Retrieved from http://www.futuristgerd.com/2013/06/02/good-read-scientists-save-healthcare-but-

[5] Harvard Business. (2014). How Big Data Impacts Healthcare by Harvard Business Review Analytics Services.

[6] Hongyong Yu, and Deshuai Wang. (2012). Research and Implementation of Massive Health Care Data Management and Analysis Based on Hadoop, Fourth International Conference on Computational and Information Sciences.

[7] Emad A Mohammed, Behrouz H Far, and Christopher Naugler. (2014). Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends, Mohammed et al. BioData Mining 2014, 7:22.

[8] Lars George.(2014). Getting Started with Big Data Architecture. Cloudera. Retrieved from URL http://blog.cloudera.com/blog/2014/09/getting-started-with-big-data-architecture/

[9] Monya Baker, et al. (2012) Functional Genomics: The changes that count. Nature and Nature Methods Vol. 482, 258-262.

[10] Brian Eastwood. (2013). 6 Big Data Analytics Use Cases for Healthcare IT. [Personnel Survey] Unpublished raw data. Retrieved from http://www.cio.com/

[11] Image of medical records. Osadcha, Juliia. Web UI Color [.png]. Retrieved from URL https://www.iconfinder.com/Juliia_Os

[12] Image of prescription, The Recycling Partnership. Recycling Extras [.png]. Retrieved from URL https://www.iconfinder.com/TheRecyclingPartnership

[13] Osadcha, Juliia. Web UI Color [.png]. Retrieved from URL https://www.iconfinder.com/Juliia_Os

[14] Image of patient. Babic, Goran. Creative Agency Bresign. Medical[.png] Retrieved from URL https://www.iconfinder.com/Bres

[15] Image of analytic representation. Retrieved from URL https://store.kde.org/usermanager/search.php?username=Sephiroth6779

[16] Image of hospital sign. Babic, Goran. Creative Agency Bresign. Medical[.png] Retrieved from URL https://www.iconfinder.com/Bres

[17] Access to Health Services. (2015, October) Retrieved from URL http://www.healthypeople.gov/2020/topics-objectives/topic/Access-to-Health-Services

[18] World Health Organization Statistical Profile. (2015), Retrieved from URL‘s: http://www.who.int/gho/countries/chn.pdf?ua=1, http://www.who.int/gho/countries/usa.pdf?ua=1