Exploration on big data analytics in healthcare

M. Thilagawathi	Dr. B. Jagadhesan
thilagawathi1995feb07@gmail.com	bjagadhesan@gmail.com
Dhanraj Baid Jain College, Chennai, Tamil Nadu	Dhanraj Baid Jain College, Chennai, Tamil Nadu

ABSTRACT

In this paper, it gives the brief information about the healthcare using the big data analytics. The healthcare industry provides the huge amount of data, those data must be recorded in the database. The reports will be generated in automatically using some specific tools. It can enhance the quality of treatment for the patients. The patients can avoid the unwanted expenses. This paper includes the benefits and challenges of healthcare

Keywords: BDA, Hadoop, Big data benefits, Healthcare, Machine learning, Smart health, Fraud analysis

1. INTRODUCTION

The healthcare industry can generate the huge amount of data, it can keep the patient record. The most data store in only in hard copy, but nowadays data must be stored in digital method rather than analog. It can be used to improve the quality of treatment. It can include the clinical data like a prescription, healthcare images, and laboratory and pharmacy detail. BDA using tools (like hadoop) it will automatically generate the reports. The healthcare industry exhales a large amount of data day by day and it can be measured in a petabyte/Exabyte scale. It can be analyzed in big data and obtain the wanted results as a report.

Size: The data can be generated in any organization or individuals, but it can increase day by day. According to the report size of medical care data in 2013 is 150 exabytes and maybe it can increase in 2,219 exabytes by 2020.

Speed: It can be referred to the velocity which the new data should be generated and moves around. The social media (like Facebook, Twitter) adds to speed as the user's views, posts.

Veracity: The main aim is to ensure the data is reliable**.** It can contain the unstructured data. The new data types include social, appliance and cell phone sources.

2. USE OF BIG DATA IN MEDICAL CARE

2.1 For personage/patients

We can analyze data like a set of people affected the same disease, and in the disease cause of the same environment or same lifestyle or any family related disease. The specific method of treatment will provide the patients. The patients got an accurate and quality treatment.

2.2 Fraud Analysis

This data could help analysis the fraud related to a medical claim. It can be used to reduce the fraud, waste, and misuse.

2.3 Safety monitoring:

The potential of transforming the healthcare organizations offer a wide range of facilities. This powerful method is used to save the patient's life.

2.4 Public Health

Public health issues must solve the analytics approach. A large amount of data help to determine needs, offer required services and predict and the future crises to benefit the population.

2.5 Proactively maintaining equipment

Hospitals and medical care have a physical and technological infrastructure. Big data allows equipment diagnostics and analytics to run in real time. Equipment performance issues can instantly toward optimal performance, and it can be preventing the superior conservation or system crash.

Advantages of healthcare
- The main advantage of big data in healthcare is detecting the disease in an easier way.
- It can be reduced to the fraud-related medical claim.
- Reduce the risk to identify the data of re-admission patients.
- No need to pay a high cost.
- Improving the outcomes by examining vitals at heal-monitoring.
- Improve the quality of healthcare.

3. CHALLENGES OF BIG DATA
3.1 Unstructured data
The healthcare industry has an unstructured data, like medical details, blogs, and update the new information, in that data increase by day-to-data. It is rick to transform the data into structured data.

3.2 Misplaced or unfinished data
Some patients can't give the personal facts or lifestyle, while they are fill-up forms. So the data stored in a digital format many fields will be empty. Sometimes it's carried the wrong value, then we can consider in the record are inaccurate or unreliable.

3.3 Quality of data
The social media has another challenge, it keeps a large amount of data and unstructured data.

4. METHODS
- The medical care/healthcare industry contains the huge data about the patient's medical histories, symptoms and treatments process are collected.
- The data mining techniques are used to doing data analytics, it can give the report and control of infection.
- The Hadoop map reduces framework is able to store the healthcare data, it is reliable. It can be used to retrieve the data easily.
- It can be included the well-designed modules such as a machine learning algorithms for analysis and survey the data.

5. TOOLS\ALGORITHMS
Big data can use the several tools like Hadoop.
Some tools are a list out here.
- Jaspersoft BI Suite
- Pentaho Business Analytics
- Karmasphere Studio and Analyst
- Talend Open Studio
- Skytree Server
- Tableau Desktop and Server
- Splunk

5.1 Hadoop
Hadoop Distributed file system is used to store the very large amount of data. The unstructured data will be converted to map reduce. HBase is a column-oriented database management system it can be run on the HDFS. it is an open source framework.

Use of Hadoop:
- It can store any kind of data in quickly.
- Computing authority
- Error tolerance
- Flexibility
- small cost
- scalability

6. PROOF/EVIDENCE
Overall goals of big data analytics in healthcare:
- A large amount of data provide the right interference to the right patient at the right time.
- It can provide special care to the patients.
- The benefit of all the components of the medical care system like-patients and medical details.

Example 1: Tradition Health Prize:
Over $30 billion people were spent for unnecessary hospital admission

Aim
- We can identify the patient's risk and which type of treatment they need.
- The algorithms will predict the days which the patients have to spend in the hospital for an upcoming year.
- It can develop the new technology to improve the healthcare and reduce the more cost to patients.

7. CONCLUSION

The big data analytics can contain the many processes, in every step, contained some challenges. The BDA can hold the bulk amount of data. The specific tools are used to store the data. It can be maintained every report for the healthcare and details about the patients. We can identify the problems and give the solution to the patients in the easiest way. BDA execute the new line of treatment. For that patients can cure earlier and save their life. The BDA using a lot of algorithm for report generating. Machine learning algorithm should be used in BDA.

8. REFERENCES

[1] Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 20thVLDB conference, pp 487–499.

[2] Ahmed S, Coenen F, Leng PH (2006) Tree-based partitioning of date for association rule mining. Knowl Inf Syst 10(3):315–331.

[3] Banerjee A, Merugu S, Dhillon I, Ghosh J (2005) Clustering with Bregman divergences. J Mach Learn Res 6:1705–1749.

[4] Bezdek JC, Chuah SK, Leep D (1986) Generalized k-nearest neighbor rules. Fuzzy Sets Syst 18(3):237–256. http://dx.doi.org/10.1016/0165-0114(86)90004-7.

[5] Bloch DA, Olshen RA, Walker MG (2002) Risk estimation for classification trees. J Comput Graph Stat11:263–288.

[6] Bonchi F, Lucchese C (2006) on condensed representations of constrained frequent patterns. Knowl Inf Syst 9(2):180–201.

[7] Breiman L (1968) Probability theory. Addison-Wesley, Reading. Republished (1991) in Classics of mathematics. SIAM, Philadelphia.

[8] Breiman L (1999) Prediction games and arcing classifiers. Neural Comput 11(7):1493–1517.

[9] Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont.

[10] Brin S, Page L (1998) The anatomy of a large-scale hypertextual web Search Engine. Comput Networks 30(1–7):107–117.

[11] Chen JR (2007) Making clustering in delay-vector space meaningful. Knowl Inf Syst 11(3):369–385.