**Big Data for HealthCare**

*Abstract:*

In this paper, we will discuss the role of Big Data in the public sector. The domain is large scale so we came to a more refine sector, i.e. public healthcare and telemedicine. Big data is still on a research scale development, so many questions remain unanswered. The limitations and precise potential of mining big data and using it in advantage of the organisation is still a quandary. Such humongous data is not easy to handle and extract.

The scaling and analysing these datasets is a key basis of competition. Chances of productive growth and innovative measures to support organizational needs becomes easier when these datasets are analysed and solutions are provided according to necessity.

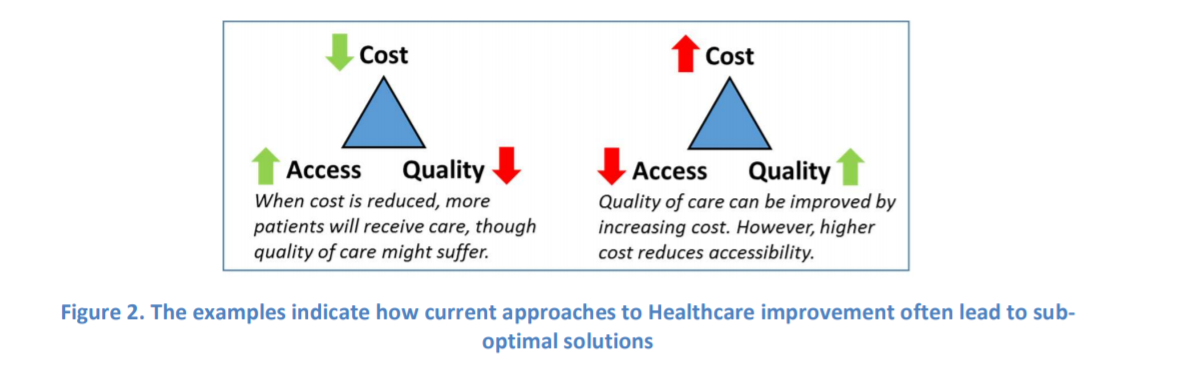
It improves the decision making processes in the hopes of solving previously ungoverned problems.

*Domain Knowledge: Healthcare*

The paper by “Big Data Technologies in Healthcare” states that the expense on healthcare is expected to increase by one third by 2060. This is mainly because of aging population, ever growing presence of chronic diseases, and costly research in development of professional medi-care solutions. Hence, by upgrading the healthcare systems, public spending on telemedicine would cut down and make considerable drop in public health care funding.

Big data technologies have already spread its effect in the public health sector starting from imaging data in medicines, to quantifying lifestyle data in the fitness industry.

It is known that the public safety sector has been trailing behind in taking big data approaches. However, it is paradoxical as 30% of electronic data storage in the world was occupied by healthcare industry, as stated by Poneman Institute in 2012. Therefore, it must have hidden knowledge that may change the life of a patient/practitioner.



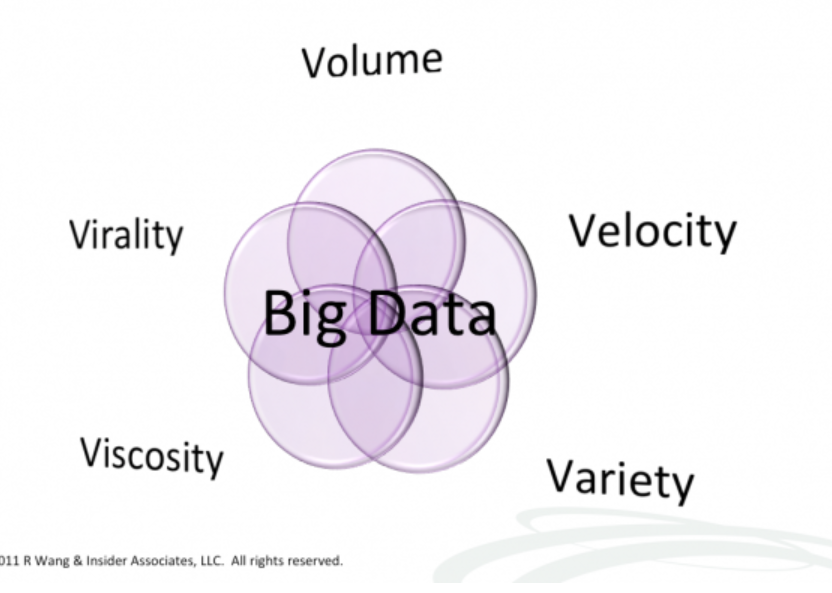
Often quoted concept of Iron Triangle of Healthcare includes quality, access and cost. While it may be possible to improve one or two of the components, however it comes at the expense of the third in most of the cases.

*Technical and Organizational Challenges:*

In the Paper” Big Data Technologies in Healthcare” big data refers to “Big data in health” encompasses high volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points. More general definition of Big Data refers to “datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse”. (McKinsey Global Institute). There are lots of issues and challenges facing with Big Data mining such that heterogeneity i.e. variety, scale i.e. volume, timeliness i.e. velocity, garbage mining and privacy

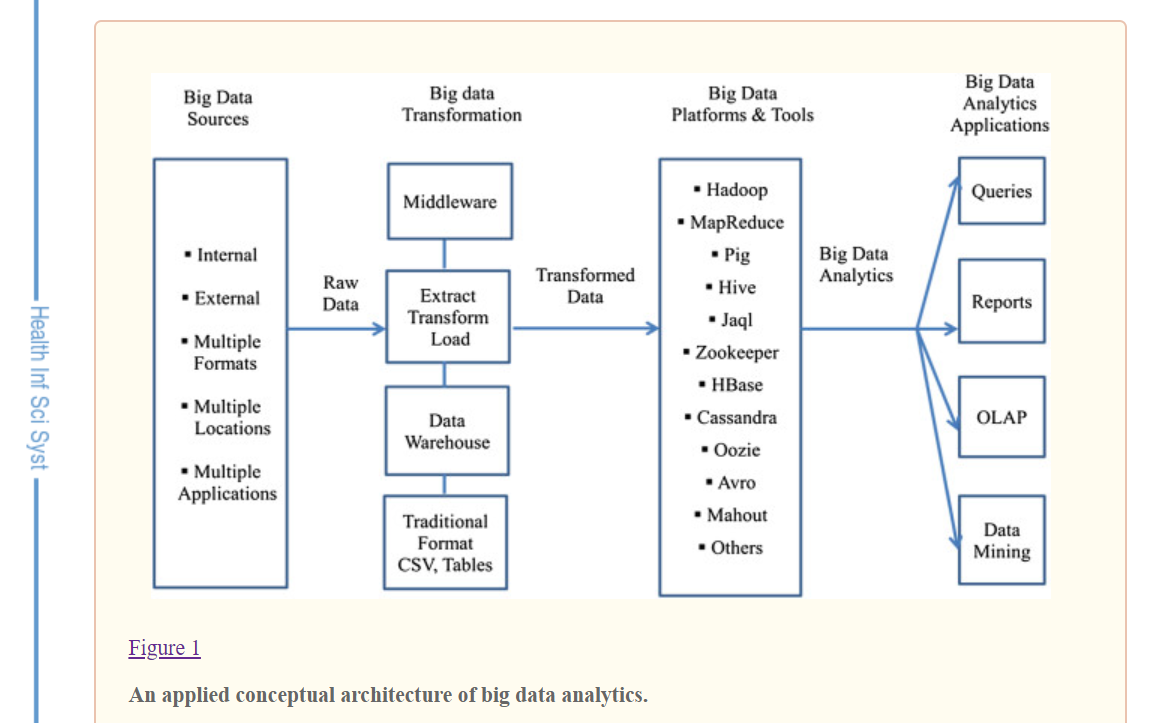
Most of the technical issues related to big data is because of scattered data. Example, we can extract the data of a patient from a hospital which exists in the EMR system, laboratory and prescription databases. Information describing the doctor and nurses attending a specific patient will also exist. However, the data source will be in separate silos. Thus, in order to derive insights, the datasets have to be merged and analysed. Any generated history of patient disease will also come into effect when calculating and deciding medicines to prescribe. It is vital to get accurate understanding of treatment trajectories of specific patients. The only way to achieve this is to be able to collect and accumulate data from different sources, not only subject to one hospital/ clinic but across multiple healthcare providers. Patient data generated from IOT gadgets such as wellness tracker, BP Monitors and gauging scales provide tactical information about daily lifestyle characteristics of a person. Unstructured data collected from the devices such as EMR data ,laboratory data, patient course of medication ,medical publications provides an insight of improving people’s lifestyle and reduces chronic diseases by thorough analysis. It is important for different stakeholders to analyse the accurate sources of data across multiple infrastructure rather than one clinic’s IT foundation. This would benefit the patients by providing right treatment , without needing to consult twice.

However, the need to integrate these unstructured datasets(granularity ,quality ,type of data) resents major lawful, business and specialized difficulties from an information point of view, as far as the volume, assortment, veracity and speed of the informational indexes. The best way to effectively address these difficulties is to use Big Data advancements.



*Data Mining Tools and Techniques*

Gartner [7] in 2012 summarizes the definition of Big data as high volume, velocity and variety information assets which demand cost-effective, information processing tools for enhanced insight and decision making. There are expansive hole between requests of the Big Data and current DBMS Storage. To beat the gap between these Hadoop was introduced which is the centre of Big Data.



The 4 v’s are the entry point to the discussion of big data analytics in healthcare. The conceptual framework of the big data is similar to traditional but relatively new is analysing the data in larger scale for better insight for making healthy related decisions

Hadoop belongs to the class “NoSQL” technologies which includes CouchDB and MongoDB.Hadoop consists Hadoop Distributed File System which is storage part and processing part which is a MapReduce programming model. This model further divide files into blocks and disperse across nodes in a cluster. This Hadoop splits files into large blocks and distributes them across nodes in a cluster with various structures or no structures at all. Since, Installing and configuration of Hadoop was a tedious task. Organizations were not able to rely on this Hadoop completely. Hence various platforms and tools were introduced that supports Hadoop distributed platform

Different Vendors AWS, Hortonworks, cloudera and MapR technologies are distributed open source Hadoop platforms. Cassandra ,HBASE ,MongoDB are database components. Other frameworks and tools are open source and these are wrapped around Hadoop and related platforms are to be considered while analysing data in healthcare by analytics developer .With the pros and cons, pros being low cost and open source and cons being less secured and lack of maintenance

In Big Data Mining lot of open source tools are available

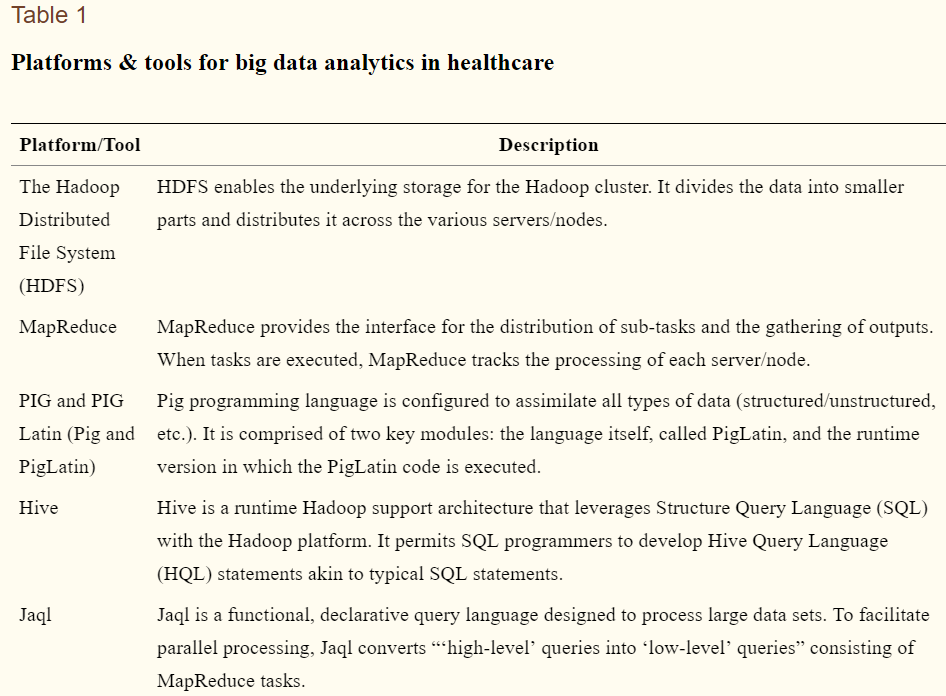
R: It is a language and environment for statistical computing and graphics. It provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering).Quality plots can be produced .

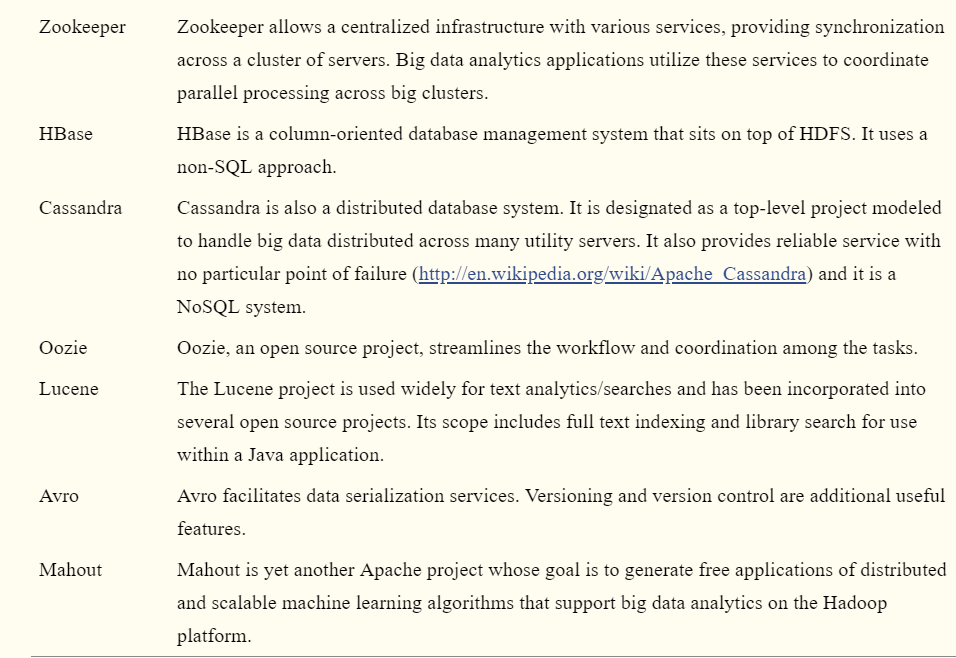
Weka : Group of machine learning algorithms to solve real world data mining problems

RapidMiner : It is an software platform developed by the RapidMiner company which provides software, solution ,and service in the field of machine learning ,Text Mining ,Business Analytics. Visualisation ,Validation and optimization of results are also supported

Knime : It is a user-friendly graphical approach for analysis process such as data access, data transformation for predictive analytics

Various other platform and tools are listed for healthcare below:

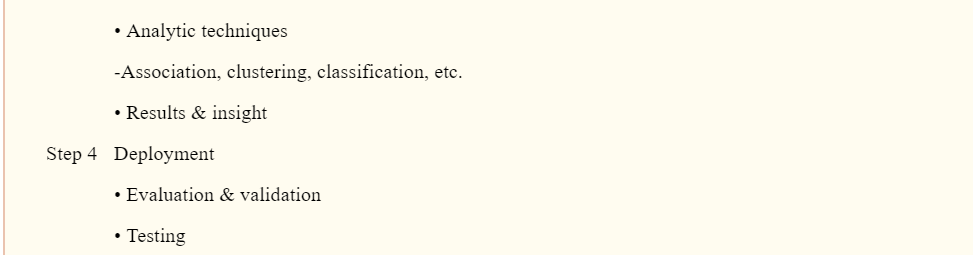
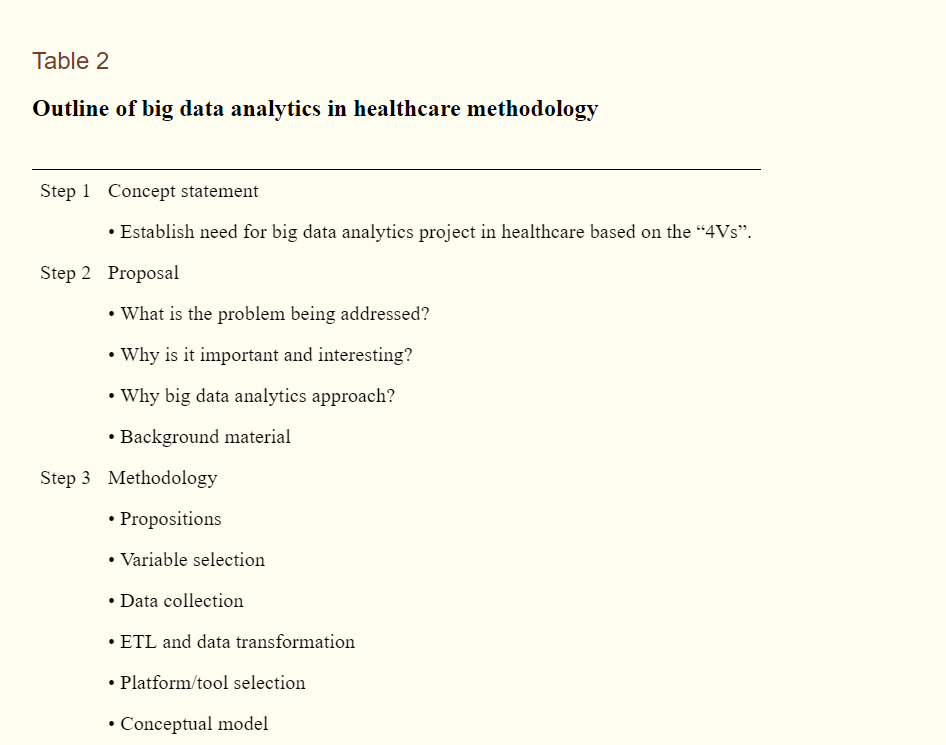




*Methodology*

Development of different methodologies .The Table 2 displays the levels of methodology.

Step1 :Big Data team develops ‘concept statement’ followed by interpretation of significance of the project



Step 2: Based on the previous feedback ,several questions are declaimed : what problem is addressed? , why it is important to the healthcare provider? , Use case of Big Data Analytics approach

Step 3: Methodologies are implemented. Data handling from various sources and transformations of data (figure 1) for analytics. Platform/ Tool Evaluation followed by application of big data analytics technique.

Techniques involved in the data mining are

Association-Searching for pattern where one occasion is associated with another occasion. It helps in discovering transactional data in medical data sets

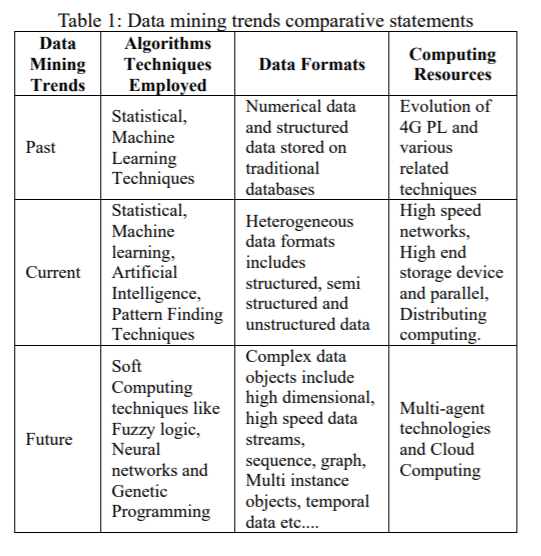
Artificial Neural Network-Non-Linear predictive model inspired by the biological neural networks that constitute animal brains

Classification : Classification is a data mining (machine learning) technique used to predict group membership that assigns items in a collection target categories or classes

Clustering : Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields. Clustering is a Machine Learning technique which includes collection of data points.

Decision Trees . It is an predictive modelling technique used in classification using divide and conquer method used in decision supporting system

Genetic Algorithm: Used to resolve problems based on natural selection based on the concept of evolution



Step 4 :Results of the models are tested and validated and presented to stakeholders.

*Success Criteria and their benefits* :

Use Case : U.S healthcare alliance network, has in excess of 2,700 individuals emergency clinics and wellbeing frameworks, 90,000 non-intense offices and 400,000 doctors and is accounted for to have information on roughly one out of four patients released from medical clinics Normally, the system has amassed an extensive database of clinical, monetary, patient, and store network information, with which the system has produced thorough and equivalent clinical result measures, asset use reports and exchange level cost information. These results have educated decision-making and improved the healthcare sector services at roughly 330 medical clinics, sparing an expected 29,000 lives and decreasing medicinal services spending by about $7 billion

*Conclusion and Future work*

We have entered an era of Big Data, By thorough investigation medicinal service suppliers adapts latest modern advances to improve their standards in the healthcare industry. Problems related to privacy, security and standards continually improve the tools which can improve their maturing process.

**References:**

Ijcsit.com. (2019). [online] Available at: <http://ijcsit.com/docs/Volume%206/vol6issue04/ijcsit20150604141.pdf>

Raghupathi, W. and Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. [online] Biomedcentral.com. Available at: https://www.biomedcentral.com/track/pdf/10.1186/2047-2501-2-3?site=hissjournal.biomedcentral.com [Accessed 10 Jun. 2018].

<http://www.bdva.eu/sites/default/files/Big%20Data%20Technologies%20in%20Healthcare.pdf>