

REVIEW OF BIG DATA ON HEALTH CARE

*T.M. Prajesh**, *S. Veni*

Abstract

Nowadays big data applications can see in almost every industry. Big data is act as a driving force behind the success of many organizations. Predictions based on big data can be applied to address issues like public health and economic development. In healthcare analytics, a set of computer-based methods are used which process raw health data and transforming it into meaningful information. This paper includes big data analysis steps to find a decision. The new discovery and its knowledge help in making more effective healthcare decisions. Several machine learning techniques can be used with big data for the better predictions in the health care.

Keywords: Big Data; Healthcare; Interpretation.

I. INTRODUCTION

In every area such as industry, business and medical have an explosive growth of data due to the rapid developments in machine learning, IOT and Cloud Computing. Nowadays big data is becoming a hot topic around the world and getting great attention from academia, industry, healthcare and government [1].

As the name mentions, Big data is very big size. Big data, the term is used to describe a collection of data which is huge in volume. Big data is growing exponentially with time. None of the traditional data management tools are able to process big data efficiently and effectively because big data is large and complex in nature [2].

There are mainly two categories of Big data. One category of data is obtained from physical world through sensors, observations and scientific experiments. Second category of data obtained from social networks, Internet, health, finance, economics and transportation areas [1]. A social media site, Facebook is getting more than 500 terabytes of new data every day which is mainly in terms of photos and video uploads, messages, comments etc..[2]. In this information era, decision makers need to handle enormous amounts of data. It is difficult to handle big data using traditional tools and techniques because they are complex and big [3].

II. CHARACTERISTICS

Characteristics of Big Data can be clearly defined by 6Vs

VOLUME: The amount of data is very large in size. Data expressed in terms of Terabytes and Peta bytes.

VELOCITY: It specifies how much data generated per second.

VALUE: Big data have huge potential value. If it can't able to turn into value, the data become useless.

VARIETY: Data come in different sources in variety of forms. Data come from different sources may be structured and unstructured. It may be in different formats such as audio, video, images, text, etc...

VERACITY: Data may contain wrong data too. Veracity specifies the accuracy. veracity of the source data captured decide the accuracy of the data.

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

VARIABILITY: It represents consistency of the data over time.



Figure1:6Vs in Big Data

III INFLUENCE OF 6VS IN HEALTH CARE

Big data solutions are used by hospitals, pharmaceutical companies and researchers to improve and advance in healthcare. In various areas such as medical research, public health, health care, etc., big data technologies have a transformative impact [4]. Effective research on diseases like cancer and Alzheimers and finding new drugs are happening due to the availability of vast amount of data. [5]. Now the healthcare analytics is popular because of the eruption of the Big Data technology in the healthcare domain [6].

A. VOLUME

In the area of medical science, health-related data will be created and collected continuously, this will result a large volume of data. Healthcare data, Personal medical records, radiology images, clinical trial data FDA submissions, etc... are included in healthcare data. Several other new forms of data are 3D imaging, genomics and biometric sensor readings. All the above medical data are also going the exponential growth [7]. When the volume of patients' data increases more accurate predictions can be done and exact

medicines can be prescribed. Several challenges occur in organizations due to the large volume of the data. One of them is that handling the significant load of information without slowing down the necessary functions like communication and access of the patients' data [8].

B. VELOCITY

In real-time data is accumulated at high speed. The flow of data at rapid rate presents some challenges. In healthcare, most of the data are in static form like paper files, X-rays and other scanning reports. Regular monitoring of patients provide large amount of data that will include daily diabetic glucose measurements blood pressure readings and heart rate reading. In healthcare, constant availability of real time data mean the difference between death and life. [7]. One of the challenges is faced due to high velocity data is that specific tools are needed for analysing, storing and managing so as to get desired output. Several forms of data remain relevant to healthcare or organizations for take the immediate actions. Sensor data is collected using health devices is one of the examples. The instantly processed health data can provide potentially life-saving information to the users and physicians [5].

C. VALUE

The value of big data is based on how it will provide actual and meaningful return on future. Patients assessment timeliness can be planned using big data. Patients with diseases like tumour and Alzheimers need periodic assessments. Using big data timeliness of the tests can be decided. This help to save time and money by scheduling tests soon.

D. VERACITY

Veracity represents big data is trusted or not. Several insights obtained from big data. If it is incomplete, it is not useful [8]. Inputs are obtained in structured and unstructured format. Increasing the integrity of data is a struggle for data

providers. All the healthcare organizations looking for standardized and complete data for better prediction [8].

E. VARIETY

Data collected exclusively in electronic health records and multimedia records. Data are collected in structured, unstructured and semi structured format. Structured data can be easily stored, recalled and processed by machine. Instrument reading is a semi-structured data. Unstructured data obtained from hospital admission records, discharge summary, hospital medical records and prescriptions of doctors [7]. Healthcare applications can be efficiently implemented, if this structured and unstructured data is converted into meaningful information.

F. VARIABILITY

Data collected from different sources may different in type. Variability represents the inconsistencies exist in data. Data loaded into database in inconsistent speed referred to as variability [9]. In health care different machines generate data in different velocity. So specific tools need to be used when people are getting sensor images of patients’ affected area in varying velocity.

IV. BIG CHALLENGES IN HEALTHCARE

There are some challenges associated with big data. They are-

A. SUDDEN GROWTH OF DATA

In Big data more and more data generated in each and every second. Data is growing at a rapid rate. It is making a challenge to process and collect information from it.

B. DATA STORAGE

It is very difficult to store and manage large amount of data. Organisations need to find out large storage spaces to hold the data.

C. SYNCHRONISATION OF DATA SOURCES

Organisations collect data from different sources. One

problem exist here is that one source might not be up to date as compared to the data from another sources.

D. SECURITY

It is very difficult to keep huge amount of data in secure. Organizations must use proper authentication and encryption methods to implement data security.

E. UNRELIABLE DATA

Big data is collected from different sources. It cannot be fully accurate. Redundant or incomplete data might be exist there. There are also chances for several data contradictions [10].

V. BIG DATA ANALYTICAL PROCESS MODEL

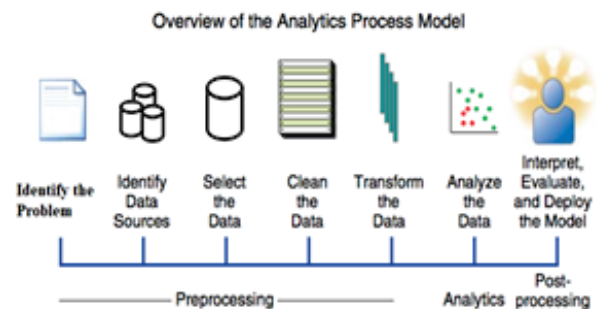


Figure2: Steps in Data Analysis

A. IDENTIFY A HEALTHCARE PROBLEM

Problem identification is the first step of analysis. Problem statement is also need to be framed. Understanding of the problem statement and the domain knowledge of the problem are required in this stage [11].

B. IDENTIFY DATA SOURCES IN HEALTHCARE

Based on problem statement, data sources need to be identified in this step. There are several healthcare data sources are available. They are Electronic Health records, patient portal, research studies and generic databases. The data collected might be large therefore use Hadoop to store and process data[12].

C. PROCESS AND CLEAN THE DATA:

Quality of data is the main concern behind the accuracy. In this step, handle missing data and transform the data to appropriate format. Accuracy of the output depends on clarity of the data. Formatting of the data is based on the selection of analytical tool [11].

D. EXPLORATORY ANALYSIS:

After completing all the above steps, now data is available for analysis. Data analysis tools are software and need to be selected to interpret and make conclusions. While performing the data analysis there exist requirement of more data, so more data might need to be collected [13]. Graphical and Quantitative techniques are available for data analysis. Scatter plots, box plots and histograms are some of the graphical techniques. Mean, median, mode and standard deviations are some of the quantitative techniques.

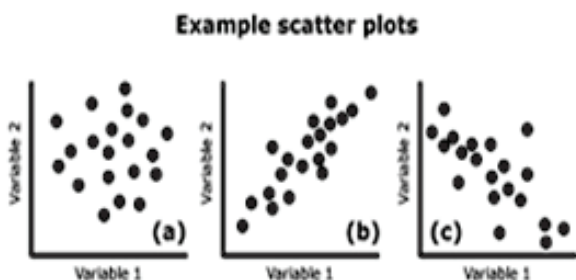


Figure 3: Graphical Techniques for data analysis

E. INTERPRETATION AND VISUALIZATION OF DATA, MODELLING

After data analysis, interpret the results. Results of data analysis can express in terms of words, tables and charts [14]. Tables and charts are the most appropriate methods for data visualization because humans can easily understand by these methods. Unknown facts are also easily obtained from the diagrams. The relationship between different variables can easily figure out from the diagrams. [14].

Once the analytical model is properly validated and approved, then it can be put into production as an analytics e.g., decision support system. While creating a model important considerations are how to represent the model output in a user-friendly way and how to make sure the analytical model can be properly monitored and back-tested on an on-going basis[12]. Model is the solution for a problem statement.

Modelling have done in the following three steps :

- 1) Selection of the appropriate model:** A model is built based on the problem defined. The Regression model is appropriate for predicting a future-event.
- 2) Training of model:** Entire data sets are divided into two. One is used for the training purpose and other is used for the testing purpose. A large portion of the data is supplied as input to model algorithms.
- 3) Evaluation of model:** Once built the model then perform testing and validation. Some portion of the dataset is used for the testing purpose[11].

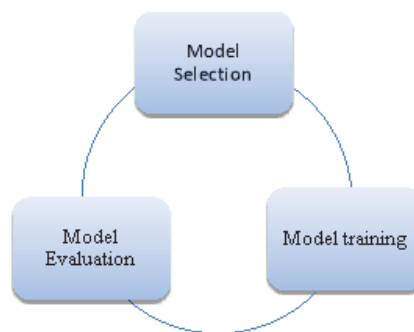


Figure 4: Data Modelling

VI. CONCLUSION

Health care providers can use big data technologies to gain insights from medical data sets and make appropriate decisions. In future, the use of big data technologies and its widespread implementations will increase. This paper has

provided the summary of steps how big data analytics can perform in healthcare. It also throws light into several challenges of big data analysis in healthcare. The big data technologies have a great impact on clinical and biotechnological research in future. Acquisitions of the data at a rapid rate make the success of big data in medical diagnosis.

REFERENCES

- [1] Xiaolong Jin , Benjamin W. Wah , Xueqi Cheng , Yuanzhuo Wang Significance and Challenges of Big Data Research Big Data Research Volume 2, Issue 2, June 2015, Pages 59-64.
- [2] Basic knowledge of Big Data, available at <https://www.guru99.com/what-is-big-data.html> on Sep 2020.
- [3] Nada Elgendy and Ahmed Elragal, “Big Data Analytics: A Literature Review Paper”, Industrial Conference on Data Mining ICDM 2014: Advances in Data Mining. Applications and Theoretical Aspects pp 214-227.
- [4] Diellza Nagavci, Mentor Hamiti, Besnik Selimi Review of Prediction of Disease Trends using Big Data Analytics (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 8, 2018 46|Page.
- [5] Characteristics of Big Data, available at <https://builtin.com/big-data> on Sep 2020.
- [6] Sherif Saki, Amal Elgammal “Towards a Comprehensive Data Analytics Framework for Smart Healthcare Services”, Big Data Research Volume 4, June 2016, Pages 44-58.
- [7] Wullianallur Raghupathi1 and Viju Raghupathi, “Big data analytics in healthcare: promise and potential” ,Raghupathi and Raghupathi Health Information Science and Systems 2014, 2:3.
- [8] Big Data on Healthcare, available at <https://www.talend.com/resources/big-data-changing-healthcare/> on Sep 2020.
- [9] Characteristics of Big Data, available at <https://tdwi.org/articles/2017/02/08/10-vs-of-bigdata.aspx> on Sep 2020.
- [10] Basics of Big data, available at <https://intellipaat.com/blog/big-data-tutorial-for-beginners>.
- [11] Data analysis steps, available at <https://www.r-bloggers.com/data-analysis-steps/> on Sep 2020.
- [12] Data analytics process model, available at <http://blogs.sas.com/content/sgf/2019/05/14/big-data-in-business-analytics-talking-about-the-analytics-process-model/> on Sep 2020.
- [13] Data analysis steps, available at https://medium.com/@kunal_gohrani/the-5-steps-of-the-data-analysis-process-2512ba6ac31e on Sep 2020.
- [14] Data analysis steps, available at <https://www.guru99.com/what-is-data-analysis.html> on Sep 2020.