

A Survey on Challenging Capabilities of Big Data Analytics in Healthcare

D.R. Krithika¹, Dr.K.Rohini²

¹Research Scholar

Department of Computer Science,

Vels Institute of Science Technology and Advanced Studies, Chennai

Email: krithikabanu@gmail.com

²Associate Professor, Department of Computer Science,

Vels Institute of Science Technology and Advanced Studies, Chennai

Email: rohini16@gmail.com

ABSTRACT: The term big data analytics is a huge volume of data which is difficult to process traditional techniques. Most of the hospitals do not maintain all necessary clinical data. Electronic health record (EHR) is a very important role to maintain a large volume of patient's data in the healthcare industry. Better outcomes for using big data analytics the Innovative Medicines Initiative (IMI) projects in healthcare. In machine learning the rules processing in edge level and healthcare cloud. The edge level is like sensor devices healthcare equipment, wearables, patient monitoring machine. It will alert the patient. Healthcare rules cloud will analysis patients sensor collected data and forecast if any health issue takes place. In this paper, we discuss big data in the healthcare industry.

Keywords: Bigdata Analytics, Healthcare, Machine Learning, EHR-Electronic health record, IMI-Innovative Medicines Initiative.

I. INTRODUCTION:

Roger Mougals from O'Reilly Media launched the term bigdata in 2005. Next year they created web2.0 refers to a large amount of data. This data is impossible to maintain traditional tools. Big data analytics examines huge and various type of data. In 2005 yahoo created Hadoop. Most of the social network start web2.0 and daily basis large created. The Indian government also decided to bigdata projects like a photograph, iris scan and fingerprint 1.2 billion inhabitants. Most of the companies slowly adapting to using bigdata. Bigdata five v's are volume, velocity, variety, veracity, value. An enormous and huge amount of data is big data. Volume is the size of a large amount of data. Velocity is the continuous flow of data and processing speed of the data. Variety is all the format of data like structured, semi-structured and unstructured. Structured data is organized data. Semi-organized data is semi-structured data. Text, video, pictures all type of data appeared in unstructured data.

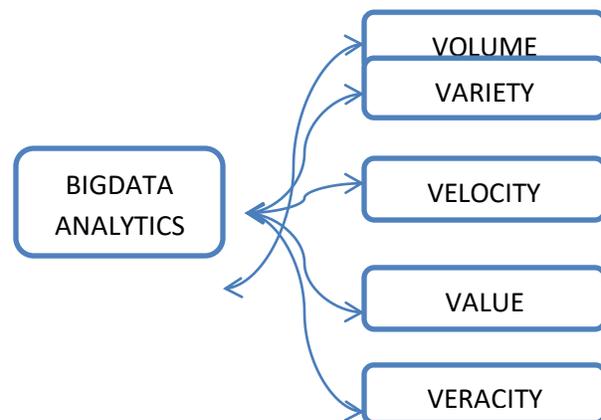


Fig.1 Five V's of Bigdata Analytics

Veracity dumping a huge amount of data some data packets bound to lose in the process, so fill up these missing data and then start mining again and process it and come up with good insight. Value is what is the analysis done it has some value, it will help to grow.

II. LITERATURE REVIEW:

The paper proposed whether the people have coronary heart diseases or not diagnosed for using linear discriminant analysis and classification algorithm [1]. The paper discusses extracting valuable information from document EHR, processed and deliver valuable both physician and patients [2]. This paper discussed World population increasing is the lead to occur various heart diseases. Most of the hospitals not vacant it will increase the death rate. In this paper resolve the things to use IoT, Nanoelectronics and bigdata. Doctors will view the patient's report from remote places. Big data and IoT combined and created medical devices. The purpose of big data is used for intelligent decision-making skills and Monitoring and controlling (sending and receiving) purpose used in IoT. This will reduce the cost [3]. This paper discusses the domain model combines ABAC (attribute-based access control) and ABGS (attribute-based group signature) and provides flexible access control [4]. In this paper discussed using deep learning algorithm to monitor and analyze patients health conditions IOT based bigdata. The IoT based medical network functions as a tool for monitoring the patient's health [5]. In this paper discovered a continuous learning process to calculate frequencies, probabilities and dependencies will form to predict process model improve economic factors and better performance [6]. The paper discussed data will be analyzed using c-means clustering algorithm and Hadoop technology to prescribe the drug to particular patients [7]. In this paper, empirical data set used to real-world ticket classification applying bigdata analytical techniques and machine learning approach to improve business process [8]. The paper proposed about medical sensors collect the patient health data and IoT thru send this data to the cloud. This IoT and cloud-based framework of big data in the healthcare domain and test the specific solution with the application of ECG monitoring and abnormality report [9]. The paper discussed BE-SYS experiment to produced result from real health data and identifying the patient at risk. Clustering algorithm and big data analytics applied in this paper [10]. This paper proposed e watch app to calculate the daily activity for obese children and adolescents [11]. This paper proposed a bigdata framework in

healthcare and the regional health information network [12]. The paper discussed digital data is a major source of the health sector and need solutions for existing issues in the health domain. The electronic health record to be sharing and retrieval of data by cloud-based [13]. In this paper reviewed healthcare data to be used in big data analytics and specifically the control of blood supply to the entire body [14]. The paper discussed the chronic disease and medical conditions are the main cause of the increase of population, cloud-based data analytics used. The cloud application layer is an interface between the users and the health network [15]. In this paper surveyed the big data analytics, Hadoop platform, NoSQL and analyzed different parameters like variety, speed, uncertainty [16]. This paper discussed digitalization[19] very huge amount of data in healthcare. The challenges of big data implementation in public hospitals[17].

III. BIGDATA IN HEALTHCARE

Tracking – In bigdata analytics the part of IoT (internet of things) tracking users statistics apart from this wearable detect patients heart rate, sleep, distance to be walked and exercised, BP etc. The sensor monitor patient's health continuously and data will be collected. The collected data identified any health issues before it goes worse. Save cost –Bigdata saves cost for hospitals, predictive analysis help to predict issues. Hospital data digitalized accessed to understand many patients. It analyzed user data and prescribed medication to avoid wrong or different medicine.

IV. ALGORITHM IN BIGDATA ANALYTICS

Linear Regression -Using these linear regression techniques to make better decisions. Simple linear regression is one explanatory variable and multiple linear regression is more than one explanatory variable. Regression analysis to prediction technique investigates the relationship between a dependent and independent variable. The data in the real world not linear, the regression lines are not intersecting or passing through the data errors which can be measured real-life distance between the data. Regression line that minimizes the errors.

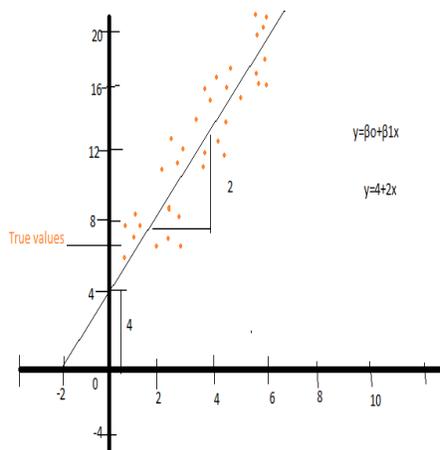


Fig 2. Linear Regression

Values X and Y to calculate linear equation:

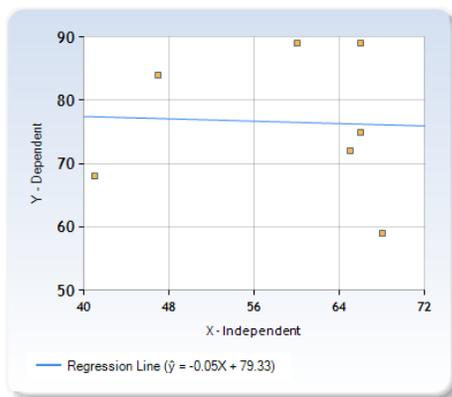


Fig 2.1 Regression Line

$$\hat{y} = -0.04678X + 79.33166$$

X	Y
41	68
68	59
60	89
65	72
47	84
66	89
66	75
M=59	M=76.5714

Logistic Regression -A statistical model is a logistic regression. Biological sciences applied this logistic regression. The model probability depends on an event occurring on values of the independent variable. The estimate is event occurs randomly selected observation. Predict the series of the variable on a binary response variable. Classify estimating observation is in a particular category. Three types of logistic regressions are binary logistic regression, multinomial logistic regressions and ordinal logistic regression.

Classification and Regression Trees- target variable is categorical and the tree gives classification. Predict the class in which the instances will fall. K-Means Clustering Algorithm[18] - The Euclidean distance is used to find out centroid is nearest to each data points assigned to the corresponding centroids. Clustering is divided into hierarchical clustering, agglomerative clustering and divisive clustering. K-Nearest Neighbors – KNN is a type of instance-based learning. The applications of KNN is text mining, advantages of KNN is robust noisy training data, training data is large, no training phase. Learns complex model easily.

V. DATA ANALYTICS

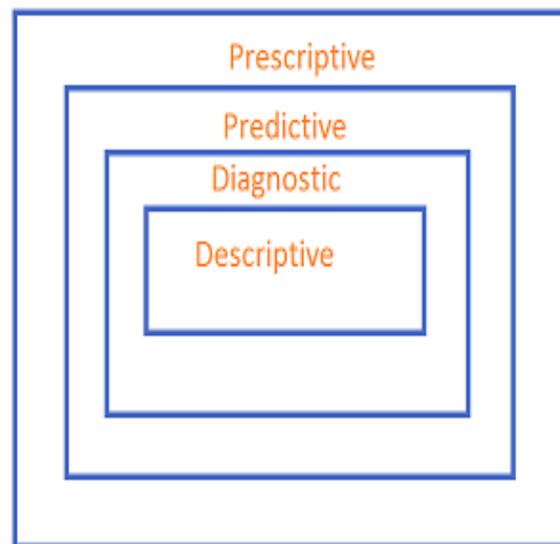


Fig 3. Different Analytics

Descriptive data analytics or observe is thinking and forming about data. Information needed for actionable decisions. Systematically collecting and interpreting data that can aid decision-makers and aware of problems. The diagnostic is the ability to know the root cause and finding the actual problems. Healthcare industry using these technologies to diagnosed the diseases.

VI. TOOLS IN BIGDATA ANALYTICS

R. Tableau, Knime, python, SAS, rapid miner, Splunk, Zoho, Aquadata studio, NodeXL, Splice machine, Spark, Apache Samoa these tools used in Data analytics. The tool tableau is a data visualization software, Tableau is easy to use drag and drop method, Any form of raw data within seconds we will result in usable data. Zoho tool is also easy to use, and quickly we get the results in the form of chart and pivot tables and required data.

VII. CONCLUSION:

In this Paper few existing research articles in bigdata, that have been implemented in the healthcare industry has been analyzed and the different data analytics to be discussed. The algorithms and tools used in existing work using big data analytics have been observed. In future work, the big data technology can be more effective to implement in the healthcare industry.

REFERENCES:

- [1] Kolukisa, Burak, Hilal Hacilar, Gokhan Goy, Mustafa Kus, Burcu Bakir-Gungor, Atilla Aral, and Vehbi Cagri Gungor. "Evaluation of classification algorithms, linear discriminant analysis and a new hybrid feature selection methodology for the diagnosis of coronary artery disease." In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2232-2238. IEEE, 2018.
- [2] Silvestri, Stefano, Angelo Esposito, Francesco Gargiulo, Mario Sicuranza, Mario Ciampi, and Giuseppe De Pietro. "A big data architecture for the extraction and analysis of EHR data." In *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642, pp. 283-288. IEEE, 2019.
- [3] Bansal, Malti, and Bani Gandhi. "IoT & Big Data in Smart Healthcare (ECG Monitoring)." In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 390-396. IEEE, 2019.
- [4] Shahraki, Ahmad Salehi, Carsten Rudolph, and Marthie Grobler. "A Dynamic Access Control Policy Model for Sharing of Healthcare Data in Multiple Domains." In *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 618-625. IEEE, 2019.
- [5] Chhowa, Tasnova Tabassum, Md Asadur Rahman, Anup Kumar Paul, and Rasel Ahmmmed. "A Narrative Analysis on Deep Learning in IoT based Medical Big Data Analysis with Future Perspectives." In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-6. IEEE, 2019.
- [6] Zayoud, Maha, Yehia Kotb, and Sorin Ionescu. "Algorithm: A New Probabilistic Process Learning Approach for Big Data in Healthcare." *IEEE Access* 7 (2019): 78842-78869.
- [7] Mishu, Md Mahub. "A Patient Oriented Framework using Big Data & C-means Clustering for Biomedical Engineering Applications." In *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 113-115. IEEE, 2019.
- [8] Lo, Dan, Karl Kevin Tiba, Sergiu Buciumas, and Frank Ziller. "An Emperical Study on Application of Big Data Analytics to Automate Service Desk Business Process." In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 670-675. IEEE, 2019.
- [9] Taher, Nada Chendeb, Imane Mallat, Nazim Agoulmine, and Nour El-Mawass. "An IoT-Cloud based solution for real-time and batch processing of big data: Application in healthcare." In *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)*, pp. 1-8. IEEE, 2019.
- [10] Martins, Francisco LJ, Joaquim Celestino, Rafael L. Gomes, Ahmed Patel, and Nazim Agoulmine. "BE-SYS: Big Data E-Health System for Analysis and Detection of Risk of Septic Shock in Adult Patients." In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1-6. IEEE, 2019.
- [11] Delopoulos, Anastasios. "Big Data Against Childhood Obesity, the BigO Project." In *2019 IEEE 32nd International Symposium on*

- Computer-Based Medical Systems (CBMS)*, pp. 64-66. IEEE, 2019.
- [12] Li, Quan, Lan Lan, Nianyin Zeng, Lei You, Jin Yin, Xiaobo Zhou, and Qun Meng. "A framework for big data governance to advance RHINS: a case study of China." *IEEE Access* 7 (2019): 50330-50338.
- [13] Gupta, Nidhi, and Bharat Gupta. "Big data interoperability in e-Health systems." In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 217-222. IEEE, 2019.
- [14] Nazir, Shah, Muhammad Nawaz Khan, Sajid Anwar, Awais Adnan, Shahla Asadi, Sara Shahzad, and Shaukat Ali. "Big Data Visualization in Cardiology—A Systematic Review and Future Directions." *IEEE Access* 7 (2019): 115945-115958.
- [15] Alamri, Abdullah. "Big data with integrated cloud computing for prediction of health conditions." In *2019 International Conference on Platform Technology and Service (PlatCon)*, pp. 1-6. IEEE, 2019.
- [16] Dhayne, Houssein, Rafiqul Haque, Rima Kilany, and Yehia Taher. "In Search of Big Medical Data Integration Solutions-A Comprehensive Survey." *IEEE Access* 7 (2019): 91265-91290.
- [17] Noonpakdee, Wasinee, Acharaphun Phothichai, and Thitiporn Khunkornsiri. "Challenges of Big Data Implementation in a Public Hospital." In *2019 28th Wireless and Optical Communications Conference (WOCC)*, pp. 1-5. IEEE, 2019.
- [18] Rohini, K., and G. Suseendran. "Aggregated K means clustering and decision tree algorithm for spirometry data." *Indian Journal of Science and Technology* 9, no. 44 (2016): 1-6.
- [19] Krithika, D. R., and K. Rohini. "Blockchain with Bigdata Analytics." In *Intelligent Computing and Innovation on Data Science*, pp. 403-409. Springer, Singapore, 2020.