IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 6, Ser. II (Nov - Dec 2019), PP 01-07
www.iosrjournals.org

A review paper on Big Data Analytics

Pallavi Ranjan, Mohita Madaan
SGT University Gurugram, India
SGT University Gurugram, India

Abstract: *In information era, enormous amount of data is produced every day, every minute and every second. This data is further analyzed to make valuable insights out of it by the analysts. Data growth is so rapid, which in turn gives birth to another concerns like managing the data properly, storing the data and maintaining the privacy and confidentiality of data. This data is further analyzed to make valuable insights out of it. So this huge amount of data is to be managed and secured. To secure our data, security challenges need to be studied. Hence this paper discusses the structured and unstructured types of data along with different stages in data management.*

Date of Submission: 11-12-2019 Date of Acceptance: 26-12-2019

I. Introduction

As the name '*Big data*' already suggests, it is a huge amount of data, having complex data set, collected from different data sources. Traditional data processing application software cannot deal with such voluminous data. So the first step is to collect the data from different sources and it is managed and processed later on. The sources where the data is collected from could be transaction systems, databases of customers, internet clickstream logs, data sensors working on real time, real-time data sensors used in internet of things (IoT) environments, data generated from machines etc[3]. Data is collected in its raw form from these sources, so using data mining tools and data preparation software it is pre-processed before it's analyzed [2].This emerging technology is facing challenges like data penetrating, data analysis, violations, sharing, data visualization, data privacy, data, security, data transition and data storage[12]. Data is categorized broadly into two types discussed below.

II. Data Categorization

Structured Data

Structured Data is consistent, easy to capture and process in the database for example in the healthcare scenario - accounting data or patient demographic data, problem list, patient medication allergy list or health history for all patients that can be fed into a database with just a radio button or a dropdown box[21]. Data can be made structured by enforcing the use of standardized data format. Structured Data is easy to upload and is fluid when it comes to analyzing the data. Some of the possible examples are:

IoT Data : Data from health monitoring devices for instance Fitbit [9] that communicates using Bluetooth to nearby mobile devices, or pacemakers and other medical devices are implanted provide data wirelessly. Stationary medical devices for bed ridden patients use wireless networks to provide data [23]

Relational Databases: Structured Data uploaded into relational databases living in fixed fields, easy to enter, extract and process. Data with patients information, transactional Data, prescriptions, research, HR, demographic data, administration or payment Records are standardized structured form of data. The OLTP database structure accommodates all kinds of data.

Lab Tests Data: Lab results become structured when fed into the electronic health systems. However, when lab experiments are performed in non-clinical settings, the results could be unstructured clinical notes[25]. In a lab experiment performed by Sijia Liu in 2018, rule based information extraction was performed to extract information from clinical notes.

Procedure Codes: American Medical Association published Current Procedural Terminology (CPT) which is a set of health care procedure codes used to define medical tests or surgeries performed on a patient. This code set is also used by insurance companies to know the reimbursement amount. Each CPD code is five digits long, could be a combination of numbers and letters. This is structured and can be easier to feed into database[25]

Unstructured Data

Human-generated unstructured data: With the increasing use of structured and unstructured data in medical records, text files, word documents, spreadsheets, emails, social media data, mobile data, digital audio, mobile data, geo-location data, photographs, and phone calls[10].To make maximum use of patient information, text-mining techniques have become extremely popular to extract information from structured and semi structured

data to make it standardized. Text Mining is extremely useful and important as it can collect unstructured data for instance, doctor's scribbles on a patient charts to analyze it for later use. Text mining is based on pattern matching, spell checking and machine learning techniques that can help to combine structured and unstructured data and ultimately lead to improved customer satisfaction[27].

Web and Social Media Data: Healthcare devices use social data to know public opinion. According to Jones (2018), social data analyses were done on the 'conversation on problems faced by the people in the US and UK' from 2015 – 2017 and it was seen that obesity is more prominent in the US conversations as compared to UK and diabetes was a more common topic of conversation in the UK. There are open health datasets available such has PatientsLikeMe, PubMed and WebMed carry a lot of structured and unstructured data.

Machine to Machine Data: M2M communication is when two machines connect without any human intervention. As mentioned in Hhmglobal.com (2018) by CISCO, there were 500 million connected devices in 2003 which then rose to 8.7 billion in 2012. It is estimated that there will be 50 billion connected devices by the year 2020.

Sensor Data: To increase customer experience, improve productivity, cost and work process, IOT devices are a primary requirement and it is believed that 40% of the total devices are health related that contributes to 117 billion market [7].

III. Big Data Challenges in Health Sector

Due to the digitization of various systems and growth in the amount of wearable technologies and the information being generated out of those technologies has increased the costs and time. To improve the overall performance of organizations or the customer experience and produce more effective data driven models, there is a need to create, organize and safeguard systems with information technology to make systems affordable, accessible and more secure[1]. Organizations analyze an immense volume and vast variety of data coming from various sources to study a clear business need, strong relationship with the customer and to have better alignment between business and IT [32].

New database management systems for instance, MongoDB, Apache Cassandra and MarkLogic allow data to travel to and fro between old and new operating systems. Big Data Analytics tools like Apache HBase and NOSQL systems nurture different varieties and huge amounts of structured and unstructured information, performing descriptive and predictive analysis on such data can result in intelligent business insights which can be used to state new business objectives[17]. A lot of research has been done so far in the area of health informatics. It is interesting to note the way health records are defined, how they are structured and used, the ethical concerns behind it, and who has access to all these records. Health records in Canada[16], England[6] and United States [34] share the common need of producing a health information system that is secure and has a common standard to access and interchange the data.

The increasing need of such a system by the citizens who want the best possible care and want to stay updated about their health regardless of their mobility [8]. This huge amount of data generated by the patient profiles, clinical data, documentation during and after delivery of care, insurance data. This data could be either structured or unstructured[19]. In 2011, 1.8 zettabytes of data was created globally, Walmart's warehouses included 2.5 petabytes of information, United States healthcare data in 2011 reached 150 exabytes, and in no time it would reach zettabytes (1021) or even yottabytes (1024). Kaiser Permanente had 26.5 petabytes and 44 petabytes of patient data (Healthcare Informatics Magazine, 2018). Big Data is extremely challenging to manage not just because of the amount of data but also the variety of data – structured or unstructured. Data could be clinical, medical reports, insurance or pharmacy or even from social media posts for instance tweets for identifying health related topics. There were 2 Billion tweets from May 2009 to October 2010 from which they try to identify potential events caused by drugs of interest[4].

Big Data Analytics has this opportunity to personalize the care by improving the quality of healthcare services and at the same time lowering costs. It can help the stakeholders provide more meaningful insights and treatments resulting in better outcomes. With information growing at an exponential rate, there is a need of Big Data systems that can identify the relationship between various sets of data, pinpoint information and improve the environment by personalizing it as per the demands of clients, combine structured and unstructured data at one place and finally a user interface with improved usability[20]. By influencing healthcare providers and by applying advanced analytics to patient behavior, best value can be provided to the customer that will lead to improved wellness. "The correlation between performance and analytics driven management has important implications to organizations, whether they are seeking growth, efficiency or competitive differentiation" [20]

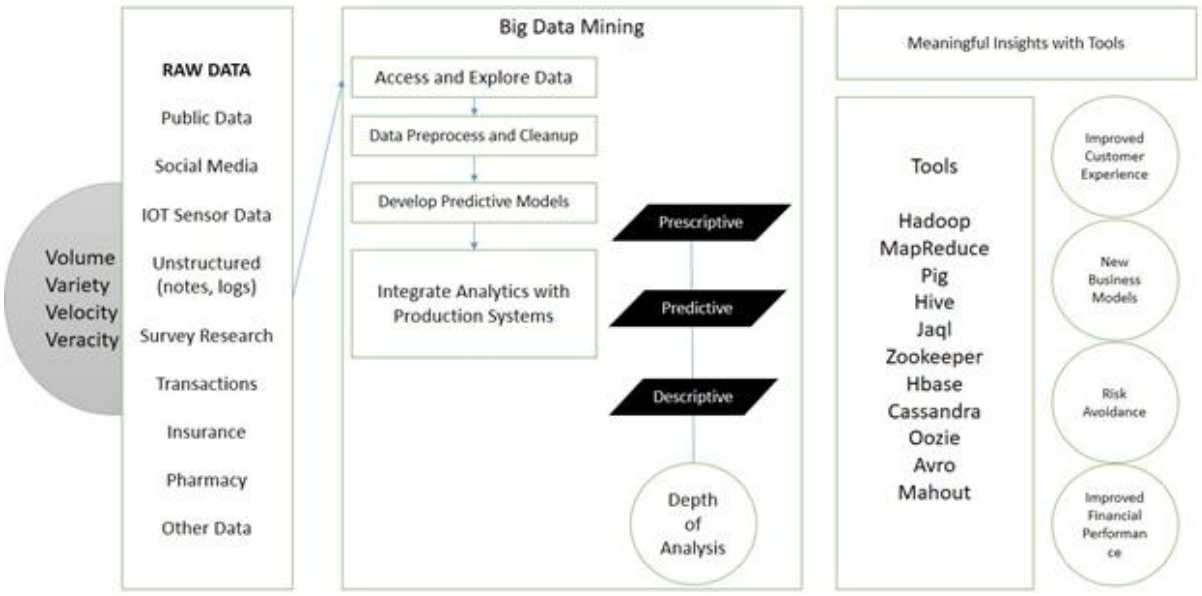

Figure 1: Different stages of Data

Huge amount of raw data for instance biological data, clinical data, patient records, prescription data, sensor data from various devices, insurance data, data from social media, biometric data or data from transactions that could be structured or unstructured is transformed with the Data mining process by extracting and reducing the complexity of data. Data is input into the big data analytics platform depending on the nature of data. After analyzing the data with the help of tools like Hadoop, MapReduce, Pig, Hive etc. Hadoop can tackle large data sets by breaking them into small partitions and later integrating them. This process can improve customer experience and new business models can be derived. This process is huge and it will be interesting to study it in more detail.

IV. Data Management

Data Acquisition: Data is generated using various IoT devices, sensors or due to interaction between devices. For instance, In a Smart Home scenario, as soon as the person arrives home, the garage door opens when the sensor senses cars, thermostat and lighting intensity is set as per the user preference. The wristband shows the steps taken and calories burned during the day. Data could be collected in active or passive mode. Active Data that is collected with the consent of the user, has a risk of annoying or losing the user and response frequency is very limited [26]. Active data is subjective to a state of person as the person's behavior may change after knowing that he is being observed. On the contrary, Passive Data is gathered without the user being involved. It runs several statistics without the user being involved actively. Passive data is objective in nature and provides in depth insights [15].

Data Validation

Data Validation is important to maintain Data Quality. It is a method of data cleansing done prior to importing and processing it. It is about the error detection and fixing of incomplete, duplicate, incorrect data by the method of replacement, modification or deletion. It is about converting the existing data set into the destination style [10]. Practically it is not possible to validate each datum as validation is an intense process.

Data Storage : Big data is a storage technology which faces variety,volume and velocity challenge. For storage purposes, a clever combination of different storage and columnar is used which is less expensive. Therefore, to accommodate all the data, distributed storage architectures are used. It handles increased storage requirements by scaling new node.So in storage cluster new nodes are being added and it is taken care of that data should be distributed between them transparently. Storage system also deals with 'velocity' because data is coming at rate and 'variety', as data is coming from different source [28]. Different writes are done on the data which slows the speed.so in that graph based database can deal with this challenge. For dealing with data explosion and scaling up and scaling down of hardware devices new big data storage systems are introduced, shifting traditional relational database systems. So the focus was to decrease query processing time but this compromised data consistency. Data warehouse is used to store structured and unstructured data. Different types of data storage:

This section assesses the current state-of-the-art in data store technologies that are capable of handling large amounts of data, and identifies data store related trends. Following are different types of storage systems:

Distributed File System-

There is a file system called Hadoop File System,is an integral part of Hadoop ecosystem technologies. It is primary data storage for many Hadoopapplications. It supports fast data transfers between all networking devices as well as it is fault tolerant. It is because each piece of data is It supports parallel processing as the content is divided into packets and these packets are distributed among nodes .So if one system crash then we have stored data on the other system[24]

No SQL database

No SQL data base model or infrastructure is developed to deal with heavy amount of data (Strohbach,M,2016).It is well adapted to do agile processing of information at a massive level. Relational databases are structured but No SQL deals with unstructured data. It follows the concept of distributed system that means the data is stored on distributed systems and servers which helps it to handle data warehouse like google, amazon etc. Distributed system concept in NoSQL makes it horizontally scalable. When the data keeps on increasing the hardware can be increased accordingly and the performance is not affected.

Columnar database

Columnar database also known as column oriented database. Columns can be added in this ·and data can be stored row wise.so when the data keeps on increasing columns can be added that makes it more adaptable.it provides scalability, flexibility. The most famous database is Hbasem.It is very good in storing data for analysis purpose later on.

Data Processing

Data Processing is about working on the data that is already stored or is produced in real time. It is about normalizing the missing data by adding values or reordering the data with the timestamps[15]. The idea behind normalizing the data is to have values set to a normal scale so that it can be utilized for future use.

Data Remanence

Data Remanence is the ability to retain the disposed data that is no longer required for future use. No removal of such data may lead to additional costs and noncompliance to privacy policies[31]. Information can be recovered using various techniques for instance, overwriting, degaussing and physical destruction. In overwriting, a pattern of zeros and ones is used to overwrite the existing data. Degaussing is a process of demagnetization where magnetic media is erased. In Physical Destruction, the media is destroyed physically so that the data cannot be recovered.

Data Analysis

After collecting data from different sources it is transformed and stored, the next step is *analyzing the data*. This is done by the following steps-

1. After collecting and transforming data,metrics are defined for a specific problem. For example identifying the potential customer who could be contacted in future.
2. Based on the timeliness of analysis, a suitable architecture is selected.As thedata keeps on changing constantly,rapid real time analysis is needed is done in that case. For instance, fraud detection in retail sectors and telecom fraud. The applications which need not work on high response time can work on offline analysis bases.Other important step of data analysis is selection of apt algorithm or technique [30].

Types of big data analysis

1. **Descriptive analytics**
   This is a technique where current and historical data is analyzed which focuses on the reason of results. Analysis is done by Descriptive statistics and inferential statistical methods[30]. Knowledge collected is summarized and knowledge patterns are represented using statistical methods, such as mean, median, mode, standard deviation, variance.Historical data is used to create management reports and identify patterns in descriptive analysis.Traditional businesses widely use descriptive analysis as it focuses on what has happened like it focuses on total stock in inventory, average,maximum or minimum dollars spent per customer and growth in sales[33].
2. **Predictive analysis**
   Predictive analytics is the use of data, statistical algorithms and machine-learning techniques to identify the likelihood of future outcomes based on historical data. Predictive models use know results to develop a

model that can be used to predict values for different or new data. The modelling results in predictions that represent a probability values for different or new data. The predictive analytics used to predict trends, improve performance, drive decision making, and predict the behavior.

3. **Prescriptive analysis**
After prediction,prescriptive analysis is done which analysis the future risks and future opportunities. The difference between predictive and prescriptive analysis is that predictive analysis concentrates on quoting probable problems that can occur in the future and prescriptive analysis focuses on suggesting the solutions and also its possible outcomes. Predictive and prescriptive analysis are dependent on each other for taking business intelligence to an another level[22].

IV. Need of Security in Big Data

Data is growing every day, every minute, every second because of which securing data has become an issue.Moreover machine (processing data) has to be validated to stay up to date. The data sources also cannot be traced and monitored performing audits on such enormous data is possible. Other reasons could be technologies that are not sufficient enough as they lack security and privacy maintenance features, Sufficient training is not there for providing security, Lack of understanding about how to handle such a large volume of data, policies are not adequate for ensuring the agreement with current approaches to security and privacy, Money is not spent by companies because of budget issues.

Security challenges with big data:

1. **MapReduce** -For big data, parallel computing and distributed framework is used. It divides the input file into different chunks and assign them to mapper which reads, processes and then gives output in the form of key/value pairs. The values are then combined in the form of key/value pairs and output is displayed accordingly.

Malicious mappers damage the data and are difficult to detect. Other issue is leaking of data by mapper, intentionally and unintentionally. The possible attack on MapReduce are Man in the Middle attack, Replay attacks and Denial of Service attack. Rogue data nodes can be added to the network that receives replicated data and can change the MapReduce code.

2. **Authentication** techniques are not strong in 'No SQL' database along with the weak password storage mechanism.NO SQL database is used to store unstructured data. They are prone to 'Man in the Middle attack' as they use HTTP authentication. And because of this they are prone to attacks like cross-site scripting, cross-site request forgery and injection attacks[29].

3. In big data, the data is stored in multi-tier storage media because of which scalability and accessibility becomes an issue. It is not easy to keep track of stored data. This is another challenge faced in **data security**.The encrypted data is stored in auto storage system and private key is distributed to each user. Users having the private key are authorized to access certain portions of data[29]. Stored data cold have security threats for instance, compromising of data servers during transmission of data that can be solved by using SSL/TLS.

4. **Validating** the collected data is a challenge because the data is collected from different software applications and hardware devices at different locations [29]. Data poisoning is another threat in which data can be changed during the transmission. Mischievous users can tamper data collecting device or can change/tamper data collecting applications that are installed on a device with the motive to enter fake or malicious data into the central data collecting system.Malicious users can create fake IDs for inputting data.Bring Your Own Device (BYOD) scenario could be followed in which malicious user can bring his own device and act like a trusted device and malicious input is data from his device into central system.These types of attacks are known as Sybil attack[29]. Malicious users can perform Man in the Middle attack like changing the temperature artificially from a temperature sensor and inputting the wrong information in the temperature collection process.The data can be changed in GPS by manipulating GPS signals.

5. **Restricting access** or visibility of data to organization,individuals or system. Existing data encryption schemes were encrypting the whole database but Raghav Toshniwal et al. (2015) proposed a new system of data encryption in which the whole database is not encrypted instead the systematic encryption is done on some parts of database to preserve some level of flexibility.·

References

[1]. Agarwal, R., Gao, G., DesRoches, C. and Jha, A. (2010). Research Commentary—The Digital Transformation of Healthcare: Current Status and the Road Ahead. Information Systems Research, [online] 21(4), pp.796-809. Available at: http://www.krannert.purdue.edu/academics/mis/workshop/papers/ra_120310_1.pdf [Accessed 2 May 2018].

[2]. Babu, M. et al. (2019). Future Trends of Business Intelligence and Big Data Analytics in Ubiquitous Environment. International Journal Of Engineering And Advanced Technology (IJEAT), 8(3s), 773-778.

[3]. Bhadani, A., Jothimani, D. (2016), Big data: Challenges, opportunities and realities, In Singh, M.K., & Kumar, D.G. (Eds.), Effective Big Data Management and Opportunities for Implementation (pp. 1-24), Pennsylvania, USA, IGI Global

[4]. Bian, J., Topaloglu, U. and Yu, F. (2012). Towards Large-scale Twitter Mining for Drug-related Adverse Events. SHB'12 : Proceedings of the 2012 ACM International Workshop on Smart Health and Wellbeing, [online] pp.25–32. Available at: http://doi.org/10.1145/2389707.2389713 [Accessed 21 May 2018].

[5]. Cdn2.hubspot.net. (2018). Data Analytics for Healthcare: Creating understanding from big data. [online] Available at: https://cdn2.hubspot.net/hub/163225/file-18212299-pdf/docs/data-analytics-for-healthcare.pdf [Accessed 21 May 2018].

[6]. Connectingforhealth.nhs.uk. (2018). NHS Connecting for Health. [online] Available at: http://www.connectingforhealth.nhs.uk/ [Accessed 21 May 2018].

[7]. Dimitrov, D. (2016). Medical Internet of Things and Big Data in Healthcare. Healthcare Informatics Research, [online] 22(3), p.156. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981575/pdf/hir-22-156.pdf [Accessed 12 Jul. 2018].

[8]. Ec.europa.eu. (2018). En 2 en Communication from the Commission on to the Council, the European Parliament, the European economic and social committee and the Committee of the regions. [online] Available at: http://ec.europa.eu/information_society/doc/qualif/health/COM_2004_0356_F_EN_ACTE.pdf [Accessed 21 May 2018].

[9]. Fitbit.com. (2019). Fitbit Official Site for Activity Trackers and More. [online] Available at: https://www.fitbit.com/in/home [Accessed 10 Dec. 2019].

[10]. Gao, J., Tao, C. and Xie, C. (2016). Big Data Validation and Quality Assurance – Issues, Challenges, and Needs. IEEE 10th IEEE International Symposium on Service-Oriented System Engineering, At Oxford, UK. [online] Available at: https://www.researchgate.net/publication/299474654_Big_Data_Validation_and_Quality_Assurance_-_Issuses_Challenges_and_Needs [Accessed 29 Oct. 2019].

[11]. Gao, J., Tao, C. and Xie, C. (2016). Big Data Validation and Quality Assurance – Issues, Challenges, and Needs. IEEE 10th IEEE International Symposium on Service-Oriented System Engineering, At Oxford, UK. [online] Available at: https://www.researchgate.net/publication/299474654_Big_Data_Validation_and_Quality_Assurance_-_Issuses_Challenges_and_Needs [Accessed 29 Oct. 2019].

[12]. Harvey, C. (2017). Big Data Technologies. Retrieved 30 November 2019, from https://www.datamation.com/big-data/structured-vs-unstructured-data.html.

[13]. Healthcare Informatics Magazine. (2018). Health IT Summit Series - Calendar of Events. [online] Available at: http://ihealthtran.com/ wordpress/2013/03/iht% C2%B2-releases -big-data-rese arch-report- download-today/ [Accessed 21 May 2018].

[14]. Hhmglobal.com. (2018). M2M communication resulting in new and exciting possibilities in healthcare. [online] Available at: http://www.hhmglobal.com/knowledge-bank/articles/m2m-communication-resulting-in-new-and-exciting-possibilities-in-healthcare [Accessed 12 Jul. 2018].

[15]. Holler, J., Tsiatsis, V., Mulligan, C., Karnouskos, S., Avesand, S. and Boyle, D. (2014). From machine-to-machine to the Internet of things. 1st ed. Oxford: Academic Press, pp.108-109.

[16]. Infoway-inforoute.ca. (2018). Digital Health in Canada | Canada Health Infoway. [online] Available at: https://www.infoway-inforoute.ca/en/ [Accessed 21 May 2018].

[17]. Jiang, P., Winkley, J., Zhao, C., Munnoch, R., Min, G. and Yang, L. (2016). An Intelligent Information Forwarder for Healthcare Big Data Systems With Distributed Wearable Sensors. IEEE Systems Journal, [online] 10(3), pp.1147-1159. Available at: https://ieeexplore.ieee.org/document/6775278/ [Accessed 3 May 2018].

[18]. Jones, B. (2018). Analysis: How Can Healthcare Services Use Social Data?. [online] Brandwatch. Available at: https://www.brandwatch.com/blog/health-care-services-social-data/ [Accessed 12 Jul. 2018].

[19]. Kudyba, S. (2016). Healthcare informatics. London: CRC Press, Taylor & Francis Group, pp.354-355.

[20]. LaValle, S., Lesser, E., Shockley, R. and Hopkins, MS. (2011) Kruschwitz N: Big data, analytics and the path from insights to value. MIT Sloan Manag Rev 2011, [online] 52, pp. 20-32. Available at http://www.ttivanguard.com/realtime/bigdata.pdf [Accessed 22 May 2018].

[21]. Mercedes Transcription. (2018). Structured vs Unstructured Data in Healthcare and Medical Transcription. [online] Available at: https://www.mercedestranscription.com/structured-vs-unstructured-data-in-healthcare-and-medical-transcription/ [Accessed 9 Jul. 2018].

[22]. Naganathan, V. (2018). Comparative Analysis of Big Data, Big Data Analytics: Challenges and Trends. International Research Journal of Engineering and Technology, 5(5), pp.1948-1965.

[23]. Priyanka, A., Parimala, M., Sudheer, K., Thippareddy, Kaluri, R., Lakshmanna, K. and Praveen Kumar Reddy, M. (2017). BIG data based on healthcare analysis using IOT devices. IOP Conference Series: Materials Science and Engineering, [online] 263, p.042059. Available at: http://iopscience.iop.org/article/10.1088/1757-899X/263/4/042059/pdf [Accessed 11 Jul. 2018].

[24]. Raghav Toshniwal, R. (2015). Big Data Security Issues and Challenges. International Journal Of Innovative Research In Advanced Engineering (IJIRAE), 2(2), 15-20. M4

[25]. Scheurwegs, E., Luyckx, K., Luyten, L., Daelemans, W. and Van den Bulcke, T. (2015). Data integration of structured and unstructured sources for assigning clinical codes to patient stays. Journal of the American Medical Informatics Association, [online] 23(e1), pp.e11-e19. Available at: https://academic.oup.com/jamia/article-abstract/23/e1/e11/2379791?redirectedFrom=fulltext [Accessed 11 Jul. 2018].

[26]. Software, E. (2019). Active data vs Passive data | Lab1 insights. [online] Lab.one. Available at: https://www.lab.one/insights/active-data-vs-passive-data [Accessed 9 Sep. 2019].

[27]. Song, M. (2018). Opinion: Text Mining in the Clinic. [online] The Scientist Magazine®. Available at: https://www.the-scientist.com/opinion/opinion-text-mining-in-the-clinic-39531 [Accessed 11 Jul. 2018].

[28]. Strohbach, M., Daubert, J., Ravkin, H. and Lischka, M. (2016). Big Data Storage. New Horizons for a Data-Driven Economy, pp.119-141.

[29]. Tarekegn, G. (2016). BIG DATA: SECURITY ISSUES, CHALLENGES AND FUTURE SCOPE. International Journal of Computer Engineering & Technology, 7(4), pp.12-24.

[30]. UthayasankarSivarajah, A. (2017). Critical analysis of Big Data challenges and analytical methods. jOURNAL OF BUSINESS RESEARCH, 70, pp.263-286.
[31]. Violino, B. (2019). The in-depth guide to data destruction. [online] CSO Online. Available at: https://www.csoonline.com/article/2130822/the-in-depth-guide-to-data-destruction.html [Accessed 3 Dec. 2019].
[32]. Wang, Y., Kung, L. and Byrd, T. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, [online] 126, pp.3-13. Available at: https://www.sciencedirect.com/science/article/pii/S0040162516000500 [Accessed 1 May 2018].
[33]. WWW.halobi.com. (2019). Descriptive, Predictive, and Prescriptive Analytics Explained. [online] Available at: https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/ [Accessed 9 Dec. 2019].
[34]. Yasnoff, W. (2004). A Consensus Action Agenda for Achieving the National Health Information Infrastructure. Journal of the American Medical Informatics Association, [online] 11(4), pp.332-338. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC436084/ [Accessed 21 May 2018].

Pallavi Ranjan. "A review paper on Big Data Analytics." IOSR Journal of Computer Engineering (IOSR-JCE) 21.6 (2019): 01-07.