ISSN (e) 2520-7393
ISSN (p) 2521-5027
Received on 18st Dec, 2019
Revised on 27th Dec, 2019
www.estirj.com

7, Vs of Big Data: A Survey

H M Sohaib Ali[1], Muhammad Jameel Arshad[2,] Irshad Ahmed Sumra[3]

Department of Computer Science Bahria University Lahore Pakistan

Email address: hmsohaibali@yahoo.com[1], jameelasmi127@gmail.com[2],isomro28@gmail.com[3]

Abstract: Big Data tremendously big data set that may be analyze computationally to expose pattern, trend, and relations, particularly connecting to human performance and connections. Big Data is an appearance used to mean a massive quantity of both controlled and amorphous data that is so outsized it is tricky to development using conventional database and software technique. Big data is used in every field of life.Ithas been said that this is age of Big Data. Big Data consist of ten Vs. In this paper, provide only a brief survey of 7 Vs of Big Data in order to understand Big Data and 7 Vs definition, its applications and merits in detail. Finally we conclude the 7 Vs uses and Big Data Utilization, as a future work for researchers and students, while moving forward.

Keywords: *Big Data, 7'Vs of big data, Big data extract and Utilization.*

1. Introduction

Big data is an area that treats methods for analyzing, systematically extracting information or otherwise dealing with data sets that are too large or complex to be dealt with by traditional data processing application software. [1] Data with multiple cases (rows) provide greater statistical power, whereas data with higher complexity (more features or columns) can give rise to higher false discovery rates. [2] Big data challenges include data capture, data storage, catalysis, search, sharing, transfer, visualization, querying, updating, information privacy, and data sources. Big data was initially connected with three key concepts: volume, diversity, and velocity [3].

When we handle big data, we cannot create a sample but just observe and see what happens. Therefore, big data often includes data with a size that exceeds the capacity of conventional common software within an acceptable time [4] and value. [5].

Fig. 1 shows of Big Data

The current use of the term big data refers to the use of predictive analytics, user behavior analytics, or some other advanced data analytics method that extracts value from data, and rarely for a particular size of data set. "There is little doubt that the amount of data now available is indeed large, but it is not the most relevant feature of this new data ecosystem." [6] Analysis of data sets may provide new correlations to prevent "business trends", diseases. And facing crime and so on. "[7] Scientists, business executives, medical practitioners, advertising and governments alike face difficulties with large data-sets in areas including Internet search, urban informatics and business informatics. Meteorology, genomics, [8] Connect omics, complex physics simulations, biology and environmental research. [9]

The given paper is divided into following sections. The Section II gives applications of Big Data. Section III through Section X discusses 7 V's and its uses, merits etc. Section XI of this paper discuss about future work. Sections XII conclude the visualization specified in this paper.

2. Applications of Bid Data

Big data applications are implemented in various fields such as banking, agriculture, chemistry, data mining, cloud computing, finance, marketing, stocks, healthcare, and more. An overview is presented specifically to project the idea of Big Data. The big data film philosophy is preserved, semi-structured and structured data, although the main focus is on unstructured data [10]. Applications of big data are as following

2.1 Big Data in Education Industry

The education industry is flood enormous amounts of data connected to students, facility, curriculum, results, and what not. Now, we have realized that proper study and analysis of this data can provide insights that can be used to improve the operational effectiveness and functioning of educational institutions. The use of Big Data and analytics is a relatively new field in higher education. The relevance of analytics is seen in depth in many areas and higher education is no different. [11]

Corresponding author Email address: *hmsohaibali@yahoo.com*

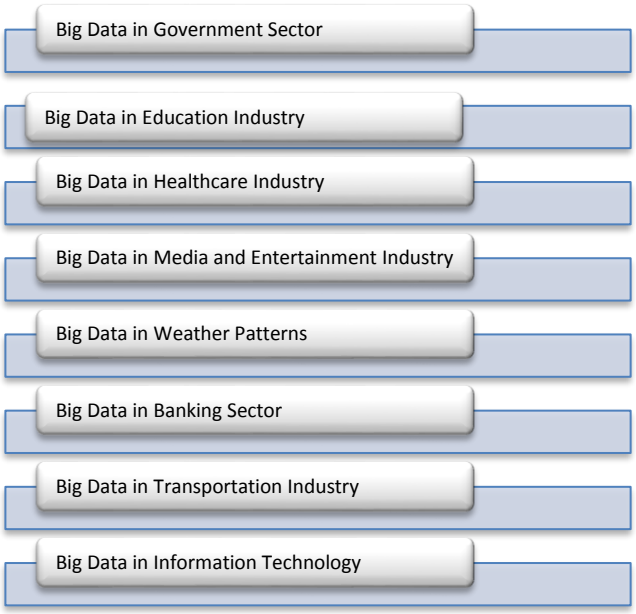

Fig. 2 shows the applications of Big Data

2.2 Big Data in Healthcare Industry

Healthcare is thus far another business vault to generate huge amounts of data. The United States has unprecedented opportunities to use big data to reduce health care costs. [12] Mobile phones, sensors, patients, hospitals, researchers, providers and organizations are generating huge amounts of healthcare data nowadays. The real challenge in health systems is how to gather information, not only to understand new diseases and treatments but also to predict outcomes in earlier stages and to make people's lives healthier and easier to make real-time decisions, be analyzed and managed. [13]

2.3 Big Data in Government Sector

Governments, irrespective of any country, face huge amounts of data on an almost daily basis. The reason for this is, they have to keep track of various records and databases about their citizens, their growth, energy resources, geographical survey and many more. Governments deal not only with the general issues of big data integration from many sources and in different formats and costs but also with some specific challenges. [14]

2.4 **Big Data in Media and Entertainment Industry**

Governments of any country face large amounts of data on an almost daily basis. The reason for this is, they have to keep track of various records and databases about their citizens, their growth, energy resources, geographical survey and many more. Governments deal with common issues of big data integration not only in many sources and in various formats and costs but also with some specific challenges.

2.5 Big Data in Weather Patterns

A weather trailing prediction consists of multiple scenarios for a weather variable. [15] Weather sensors and satellites are deployed worldwide. Large amounts of data are collected from them, and this data is then used to monitor the weather and environmental conditions.

All data collected from these sensors and satellites contribute to big data and can be used in various ways such as:

- Weather Forecast
- To study global warming
- Understand the pattern of natural disasters
- Make necessary preparations in case of crises
- To estimate the availability of usable water worldwide

3. 7, Vs of Bid Data

The seven V's summation it up appealing able-bodied, Volume, Velocity, Variety, Variability, Veracity, Visualizat ion, and Value.

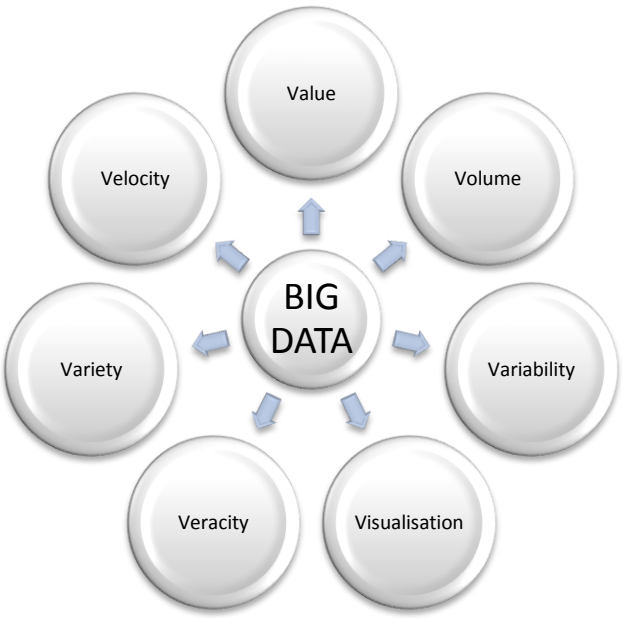

Fig. 3 shows the 7, Vs of big data.

3.1 **Volume**

Big data consistently encompasses high-volume data sets, including the ability to regularly use software tools with high volumes, to set, curate, handle, and process data within a bearable time. Is beyond A huge amount of data sets are created every second from every part of the world, i.e. the amount of data can never be reduced but increases day by day. About five years ago, personal computer storage was tens to hundreds of gigabytes. Today IDC's Digital Universe Study predicts that digital information data will increase from 0.8 ZB to 35 ZB between 2009 and 2020. Many surveys expect the amount of data to increase by 45% over the next two years, and some said it would double. Thus, big data is a dynamic target and needs more focus on capturing, curetting it, handling it, and processing it.

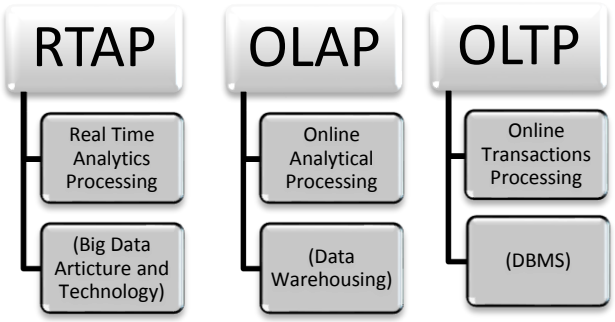

Fig. 4 shows the exponential growth of big data volume with time

The internet-mobile revolution, bringing it with a raft of social media updates, an explosion of censored data and e-commerce from devices, means every industry is swept away with data - which can be incredibly valuable Is, if you know how to use it.

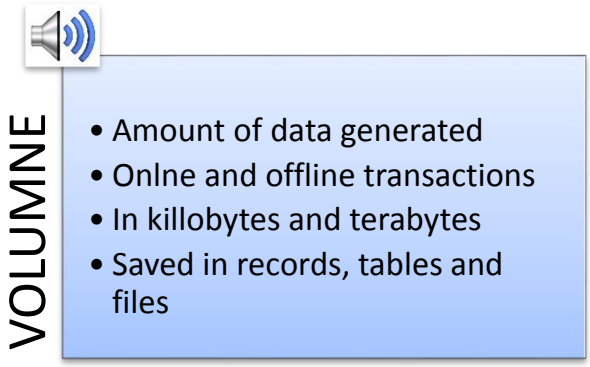

Fig. 5 shows volume of Big Data

3.2 Velocity

In 1999, The Wal-Mart's data warehouse store 1,000 terabytes (1,000,000 gigabytes) of data. In 2012, it had access to 2.5 peta bytes (2,500,000 gigabytes) of data. [11] Every minute of every day, we upload 100 hours of video to YouTube, send over 200 million emails and send 300,000 tweets. This velocity 'refers to the increasing speed at which this data is created, and the increasing speed at which data can be processed, stored, and analyzed by relational databases. The possibilities of processing data in real time is an area of special interest, which allows companies to do things like display personal ads on the web pages you visit based on your recent search, viewing and purchase history.

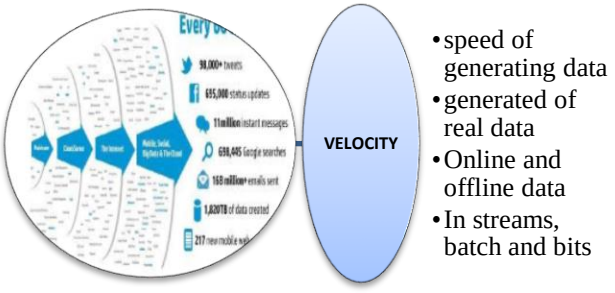

Fig. 6 shows velocity of Big Data

3.3 Variety

Departed is the existence when a company's data can be neatly placed on the table and analyzed. 90% of the data generated is 'unstructured', which comes in all shapes and forms — from geospatial data, to tweets that can be analyzed for content and emotion, to visual data such as photos and videos.

The Into 3 V 'certainly gives us an insight into the almost indirect scale of the data, and the break-neck speed at which these huge datasets grow and multiply. But only Variety really starts scratching the surface of depth — and importantly, the challenges of Big Data. A 2013 article by Mark van Rijmenam proposes Char and V to understand the incredibly complex nature of Big Data.

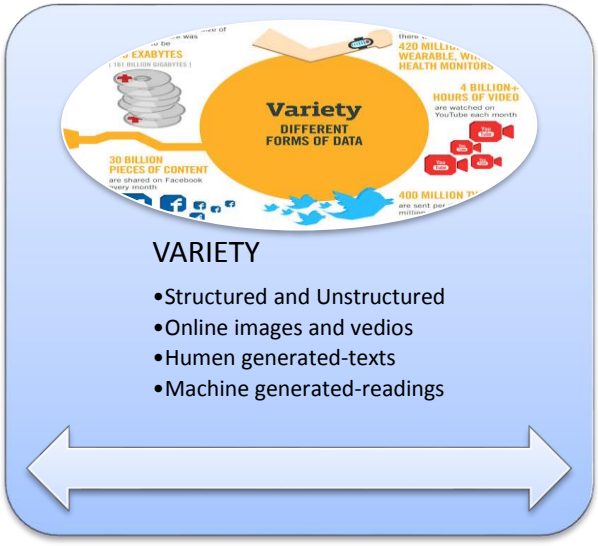

Fig. 7 shows variety of Big Data

3.4 Variability

Variation refers to data that is constantly changing. This is particularly the case when collecting data depends on language processing. Brian Hopkins (a leading analyst at Forrester) cited a prime example of supercomputer Watson. To run into the danger of the game show, Watson had to "set aside an answer in its own sense and [...] find out what the right question was". Words do not have a stable definition, and their meaning can vary wildly in context. Says that a company was trying to dispel the feeling of a cafe using these 'tweets': "Delicious Muesli from @imaginary cafe - a great way to start the day!" "Very disappointed that my Local Imaginary Cafe has stopped stocking BLT."

"Had to wait 45 minutes in line at Imaginary Cafe today. Great, well there my lunch break ... "Obviously," great "is not a sufficient signatory of positive emotion in itself. Instead, companies must develop sophisticated programs that understand the context and through Decode the exact meaning of the words from the. Although challenging, it is not impossible; for example, Bloomberg spent the last year thinking about companies for Wall Street. L'media discussion, start a program.

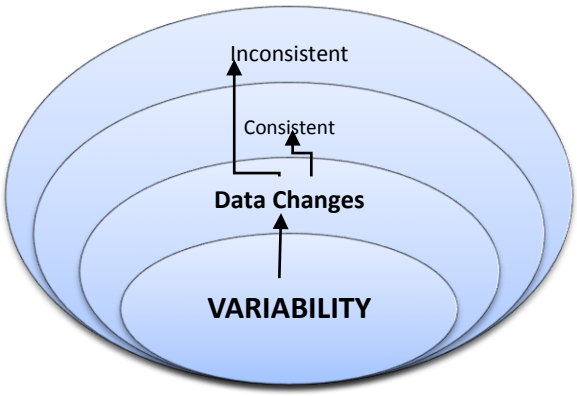

Fig. 8 shows variability of Big Data

Variability of Big Data represent data flow can be consistent with regular peaks or inconsistent. Data flow is changeable and it has two parts i.e. consistent and inconsistent.

3.5 Veracity

Big data variance refers to bias, noise, and abnormality in data. Whether the data that is being collected and the problem being analyzed, has meaning. The biggest challenge is to make Inderal feel the truth in data analysis as compared to things like volume and velocity. Unfortunately, sometimes instability is not under our control. Volatility, sometimes called another "V" of big data, is the rate of change and lifetime of data. An example of highly volatile data includes social media, where emotions and trending topics change quickly and often. Less volatile data will look more like weather trends that change less and are easier to forecast and track. While there is a broad consensus about the potential value of Big Data, data is virtually useless if it is not accurate. This is especially true in programs that involve automated decision making, or feed data into unheard machine learning algorithms. Such programmed results are only as good as the data they work with. Sean Owen, senior director of data science at Cloud On, elaborated on this: Let's say that, in theory, you have customer behavior data and want to predict purchase intent. In practice, you have log files from six systems in four formats, some incomplete, with noise and errors. These have to be copied, translated and integrated. What is important for understanding Big Data is the muddled, noisy nature of it, and the amount of work that goes into producing an accurate dataset before analysis also goes.

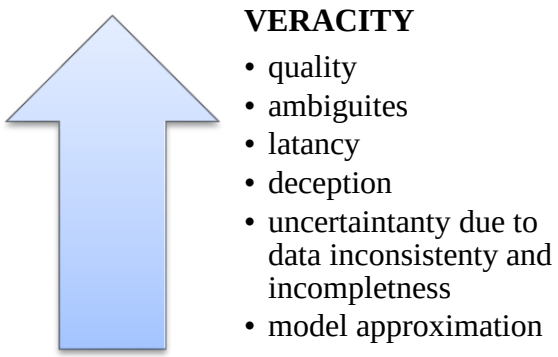

Fig. 9 shows veracity of Big Data

3.6 Visualization

Big data visualization allows a comprehensive visual representation of data patterns, enabling users to shine through data more easily. Big data visualization refers to the implementation of more contemporary visualization techniques to illustrate relationships within data. The visualization strategy includes applications that can display real-time changes and more illustrative graphics, thus going beyond pie, bars, and other charts. Once it is processed, you need to present the data in a way that is readable and accessible — this is where visualization comes from. Visualization can have dozens of variables and parameters — your standard bars cry far away from the x and y variables. Charts — and finding a way to present this information that clarifies the findings — is one of the challenges of Big Data. This is a problem that leads to a growing market — new visualization packages are appearing throughout, with AT&T announcing its offering, Naan Cubes, just this week.

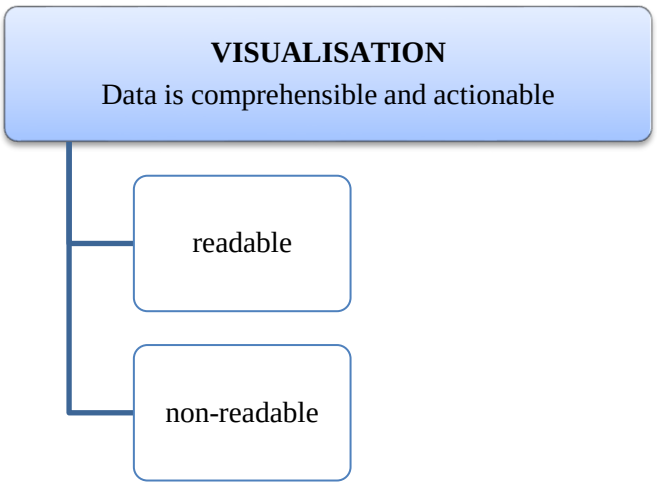

Fig. 10 shows visualization of Big Data

3.7 **Value**

When we speak about value, we pass on to the value of the data being extracted. It is one thing to have endless amounts of data, but it is useless until it can be converted into value. While there is a clear link between data and insights, it does not mean that Big Data has value. The potential value of Big Data is huge. Talking about new Big Data initiatives in the American healthcare system last year, McKinsey estimated that if these initiatives were rolled out system-wide, they would "spend $ 300 billion to $ 450 billion in healthcare spending." Or $ 12 to 17 percent to $ 2.6. Trillion baselines in US health care costs. "However, the cost of bad data is also very large — it is estimated that American business the cost of SAY is $ 3.1 trillion per year. In short, data on its own is virtually worthless. The value lies in rigorous analysis of accurate data, and it provides information and information.

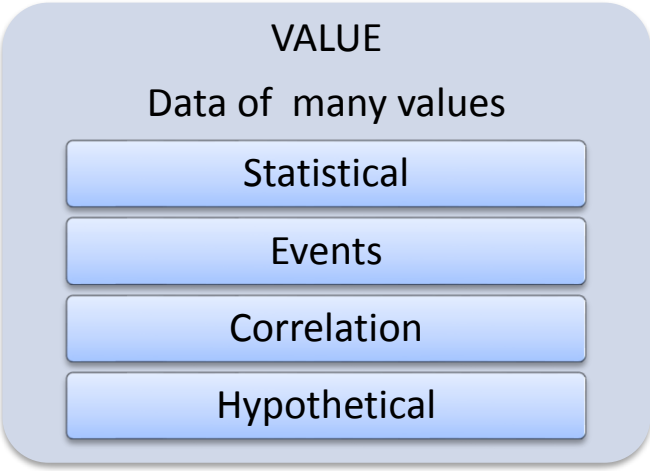

Fig. 11 shows value of Big Data

Added value that the collected data can brings.

It refers to the value that the data adds to creating knowledge.

7vs	Examples
Volume	• Social media platforms • business processes, • machines, • networks, • human interactions, • Data warehouses.
Velocity	• Number of emails, • twitter messages, photos, video clips, etc.
Variety	• Estimated 80% of all enterprise data is unstructured • Open Data(Government: https://www.data.gov/open-gov/) • NoSQL Databases
Variability	• Big data is also variable due to the multitude of data dimensions generated from many heterogeneous data types and sources. • Variability can also refer to the inconsistent speed at which big data is loaded into your database.
Veracity	• Consider a set of statistics on what people buy at restaurants and the prices of these items in the last five years. You may ask: Who created the source? What method did they follow in collecting data? • Answers to these questions are necessary to determine the veracity of this information. Knowledge of the veracity of data in turn helps us to better understand the risks associated with analysis and business decisions based on this particular data set.
Visualization	Visualization is important in today's world. The use of charts and graphs to visualize large numbers of complex data is much more effective at expressing meaning than a spreadsheet and is filled with numbers and formulas.
Value	Significant value can be found in big data, including better understanding your data, targeting them accordingly, optimizing processes, and improving machine or business performance. Before adopting a big data strategy you need to understand capacity with more challenging features.

Table.1 Shows 7 Vs with Examples

4. Future Work

After a study of big data and a line of current state of the art situations and then trivializing the ideas in this paper, it put future work at the bottom point. Designing specialized tools and techniques to extract value from such large stream data that can be used for a specific industry. Developing very specific algorithms to use 6 vs. Big Data to find 7V.Developing healthcare solutions to use big data to improve health care, disease control, early diagnosis and drug production. An algorithm should be written to develop evidence base medicine and personalized medicine using

existing scientific evidence and trial data. Cognitive science, AI to explore availability of education in rural areas, promote renewable energy and clean environment, implement safeguards to keep everyone safe, improve industry and economy and employ more people And to develop machine intelligence using Big Data. Working around the world, controlling crimes and solving poverty issues in developing countries.

5. Conclusion

Big data is a compilation of records set that are growing day by day as data is created for everyone and everything from mobile devices and call centers. This paper revolves around big data and its features such as volume, velocity, value, diversity, accuracy, diversity, visualization of 7V. The day-to-day reports show that the available data is not sufficient to manage and process big data. Hence the research presented in the paper has explored big data to identify seven big vs. its properties and big data applications.

References

[1] Hilbert, Martin; lópez, Priscila (2011). "The world's technological capacity to store, communicate and computer information". Science. 332 (6025).

[2] Breur, tom (July 2016). "Statistical power analysis and the contemporary "crisis" in social sciences". Journal of marketing analytics. 4 (2–3):

[3] Ianey, Doug (2001). "3d data management: controlling data volume, velocity and variety". Meta group research note. 6 (70).

[4] Li, Rita; li, hero (29 January 2018). "Have housing prices gone with the smelly wind? Big data analysis on landfill in Hong Kong". Sustainability. Mdpi Ag. 10 (2): 341.

[5] Marr, Bernard (6 march 2014). "Big data: the 5 vs. everyone must know".

[6] Boyd, Dana; Crawford, Kate (21 September 2011). "six provocations for big data". Social science research network: a decade in internet time: symposium on the dynamics of the internet and society. .

[7] Jump up to: "data, data everywhere". The economist. 25 February 2010. Retrieved 9 December 2012.

[8] "Communityclevernessrequired". Nature. 455 (7209): 1. September 2008. .

[9] Reichman OJ, Jones MB, Schildhauer mp (February 2011). "Challenges and opportunities of open data in ecology". Science. 331 (6018).

[10] *Dedić, n.; stanier, c. (2017). "Towards differentiating business intelligence, big data, data analytics and knowledge discovery". Innovations in enterprise information systems management and engineering. Lecture notes in business information processing. 285. Berlin; Heidelberg: Springer international publishing. Pp. 114–122*

[11] B.tulasi, (assistant professor department of computer science Christ university bangalore,india), "significance of big data and analytics in higher education", international journal of computer applications (0975 – 8887) volume 68– no.14, April 2013 21

[12] Bate set al, "big data in health care: using analytics to identify and manage high-risk and high-cost patients" health affairsvol. 33, no. 7, July 2014

[13] Hiba asri, Hajarmousannif, Hassan al Moatassime, Thomas noel, "big data in healthcare: challenges and opportunities", cloud technologies and applications (cloudlet) 2015 international conference on, pp. 1-7, 2015.

[14] Gang-hoon Kim, sylvan trim, and ji-hyong chug. "big-data applications in the government sector", communications of the ACM, March 2014, vol. 57, no. 3.

[15] J. W. Taylor, r. Buizza, "neural network load forecasting with weather ensemble predictions", *power systems IEEE transactions on*, vol. 17, no. 3, pp. 626-632, 2002.

[16] Arockiapanimalar, varnish shree, veneshia Katharine (sept 2017). "The 17 vs. of big data". International research journal of engineering and technology volume: 04 issue: 09.