

Role of Big Data Analytics in Power System Application

Ravi V Angadi^{1*}, P. S Venkataramu², and Suresh Babu Daram³

¹Dept. of EEE, SoE, Presidency University, Bengaluru, India

²Dept. of EEE, SoE, Presidency University, Bengaluru, India

³Dept. of EEE, Sree Vidyanikethan Engineering College, Tirupati, India

Abstract. Power system sector is the back bone for any country economic growth. In current years, electric power systems have experienced various challenges and technological innovations and have become digitalized with the introduction concept of smart grids. Power systems are being operated in a stressed condition mainly due to the ever increasing load demand, depleting energy resources and environmental constraints on Transmission line expansion. This article focus mainly role of Big Data in various industrialization in brief and specifically applied in the power system studies along with other sectors. Also focuses on using very large data collections, which are difficult to access in standard database systems and also refers to as big data, to manage and monitor the power system. System stability is an significant goal for power engineers to use this huge amount of data to run the system in their rated capacity, power sector can beneficial of various potential solicitation of power system by the use of large-scale data analysis that can help improve the optimization process and helps for the power system to operate in the effective manner.

1 Introduction

India is the world's 3rd largest electricity producer and consumer. India has a surplus amount of generation capacity but lack of proper corridor to distribute the generated power. To improve the corridor to distribute the generated electricity to the consumer Govt. of India is launched a program called "Power for all" in the year 2016. This program main agenda is to providing the necessary infrastructure to ensure the uninterrupted power supply to domestic, industrial and commercial establishments. Also the funding was provided with collaboration between the Govt. of India and its constitutes [1]. While many of us agree the electricity flow when we turn or connect our chargers, a broad and complicated network is needed to ensure our energy reliability.

The network connects electrical producers and consumers through a complex transmission and distribution network which connects almost all buildings across the country. Power generators use a range of turbine systems, from coal to natural gas to country-wide facilities. Electric power is transmitted from the generating station to the transmission network, where appropriate. The transmission grid consists of high voltage lines linking the generators to distribution intersections [2]. The system's reliability allows for the power to flow most places even though an obstruction occurs in the line or generators collapse suddenly and the voltage is decreased and transmitted to end users, particularly at distribution points.

To ensure the smooth running of the grid, there are many considerations. The goal of the producers of electricity

is to generate and sell energy as expensive as possible. Their principal duties consist of repairing and modifying generation equipment on the basis of input from an equilibrator. The balance body is an impartial agency that guarantees that the transmitting network is strong enough to satisfy the consumer's demand, but not an unreasonable one. They will order power generators to adjust the power output on the basis of the whole network on the basis of real-time condition, taking account not only of demand, but also of considerations such as network transmission capability on different lines. In addition to these variables, the network can also be dynamically reset, opening and closing switches.

The utilities must manage the delivery system to ensure the customer has electricity. A utility may also actively reconfigure the energy flows in response to both expected and unplanned actions within its delivery network. Critical coordination between positions is necessary, even under regular operating circumstances under order to ensure grid stability.

This article consists of the different parts, in the second section it offers an overview and Collection of big data applications in various sectors. Within the third section, the Big Data role in power systems is defined, along with various sources and features of Big Data, within section IV the Big Data challenge in the power system is discussed.

2 Big Data Analytics in Various sectors

Big data analysis continues to power daily life and the focus has significantly changed from the technologies surrounding it to the quest for practical significance. Within this time, the data analysis is a big step. While

* Corresponding author: raviangadi4045@gmail.com

the use of Big Data is still a challenge, other practical limitations such as financing and returning resources and abilities continue to lead to a large number of different Big Data industries. According to research and industry reports, 2017 saw the global big data market as \$35 billion and \$157 billion by 2027 [3]. Some of the priorities to implement or adapt Big Data programs in general. Although most organizations' primary objectives are to optimize customer satisfaction, cost savings, better collaboration and the efficiency of internal processes are other targets. The below figure 1 shows Big data growth in Billions of Dollars industry revenue



Fig. 1. Big data growth in Billions of Dollars

Even more specifically, where are we in relation to the big data? We'll maybe say that we are;

- Seek to determine whether the big data is important or not.
- Determine the size of company order.
- Innovative technology and products based on results.
- There are already large data applications in use.

Upload the existing infrastructure and big data goods. The top ten big data industries face particular problems in these sectors and how big data systematically addresses the issues that have been found. In the same top ten industries, data solutions are shown in the below figure 2.



Fig. 2. Role of Big data in various applications.

Out of all these top ten potential applications, Big Data is an important field for the electricity industry. Smart meters are also used for calculating energy consumption in the electricity industry. These intelligent meter readers can collect data almost every fifteen minutes and use these data. To improve the analysis of utility usage periodically to allow improved and better monitoring of utility usage. In the following next sections, any way the

focus of attention is to applying of big data to the power systems.

3 Big Data Role in Power System

The electricity system has faced numerous challenges and technical advances. To operate, maintain and regulate the electric power system from the generating sector, we need to tackle these challenges and technological issues effectively. We need to conserve and analyse the past data which are obtained in a variety of phases, by using the transmission and distribution part of this process.

Big data analysis should be more suitable for the power system market. The measuring data in electrical transmission and distribution systems are growing with a large number of developments in monitoring and sensor network the range and speed. The focus of the real time power generation, transmission, distribution, and use optimisations of the Big Data analytics in power systems, together with precise load demand forecasting, use patterns, new service and dynamic tariffing for the various consumers are also needed.

3.1 Sources of Big Data in Power System.

The most popular belief is that Big Data is about the data size or volume. But in fact, isn't just the large amount of information that is collected. Big Data refers to the large quantity and format of data which comes in the form of various data sources. Even before, huge information was stored in the database, but the traditional database systems are unable to handle such data because of the diversity of this data. Big data is much more than a series of data sets in various types, it can be used as a significant tool to achieve the system's multiple advantages.

The smart grid of the modern electricity system that makes two-way energy flows via two-way communication and control capabilities leading to a range of new functions and applications shown in Figure 3 below. Now the modern era smart grid plays an important role to meet the customer needs and utilities sectors. A centralized unit is collect the data by using the Advance Measuring Infrastructure such as smart meter and other specific devices collects the huge data and sent back to the centralized unit vice versa. Various measuring devices are installed at the different stages based on the requirements, these all devices are lead to huge data and which are in bidirectional and flows to centralizing unit and consumer premises this results in huge volume of data in different forms.

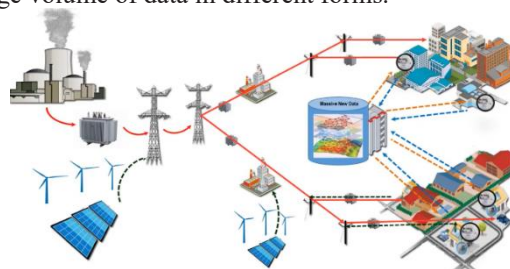


Fig. 3. Power System Network with a Centralized Unit collecting massive data from various sources.

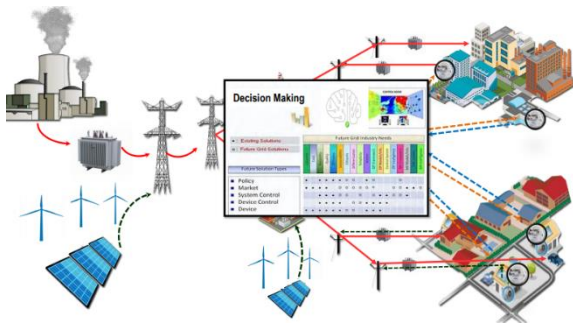


Fig. 4. Power System Network with a Centralized Unit along with decision making system.

The huge data is generated and collected in the centralized unit, this data will be in the different forms, so in order to extract the important data from this huge chunk of data we need to go with a advance competing techniques because these data are in Complexity Dimensions, so we need to use optimization algorithm and in some case there is chance of data is missing or data will be uncertainty so data cleaning is required. Once this is process is done, we can go for the decision making by using the advance Machine learning & AI techniques for the policy making, marketing, system and device control for the better and operation and maintenance of power system. The figure 4 shows the power system network with a centralized data unit and decision making operation.

The main Big Data sources in power System are as follows;

- a. Phasor Measurement Unit (PMU).
- b. SCADA.
- c. Smart Meters (SM).
- d. Updates of Weather.
- e. Geographic Information System (GIS).
- f. Simulation and User Inputs.
- g. Social Media.
- h. Traffic updates.
- i. Remote Terminal Unit (RTUs).
- j. Programmable thermostat and many more.

3.2 Big Data Characteristics in Power System.

In the above section we have seen about the how the data is generated and the various devices that are generates large set of data. Each device will generates the data in the form of different forms, based on this data are characterised into four important V's.

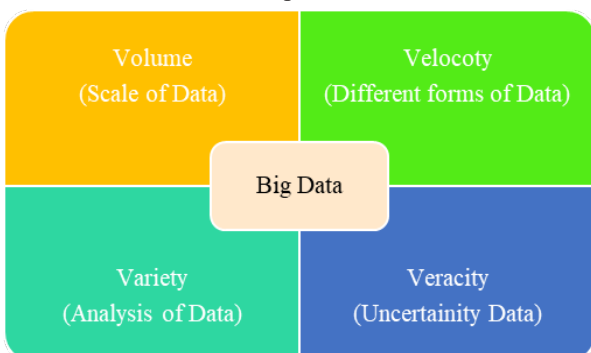


Fig. 5. Characteristics of Big Data

The fact is that big data meetings are nowadays common for organizations. Big data refers to the use of various technologies, both past and new, to extract important information from huge chunk of data. The data set is large and also has a special collection of Capture, control and process challenges. Big data format can be structured, semi structured to understand, or collected with the different resources in different sizes, in contrast to the information in relation databases structured.

We need to look back at important characteristics of large-scale data and how these features provide knowledge about data in order to understand the big data. More than simply a huge data collection, big data is described. The broad data has essential characteristics of 4V [4]. The above figure 5 shows characteristics of Big Data. These 4Vs apply to various systems that are available for use by large-scale data uses [5-19]

i. Volume: This feature refers to the total volume of the ever explosive data in the computing world. It raises the issue of the volume of data. The large amount of data from the intelligent metric and sensor technology is used in the intelligent grid to collect and store records in the register. Table I below describes periodically the volume of the data class and the source;

Table. I: Data class and Data Source of Volume.

SL. No	Data Class	Data Source	Volume
1	Utility Measurements	Phasor Measurement Unit (PMU)	30GB per day
		Smart Meter (SM)	120 GB /day
2	Weather Data	Radar	612MB/day per radar Scan
		Satellite	At least 10GB per Day
		Weather Forecast Model (WFM)	5-10 GB/day per Model
3	Vegetation and Topography	Ecological Mapping System of Texas	2.7 GB per day

ii. Velocity: This feature refers to the processing speed necessary for the collection, processing and use of large data in power systems. The data collection time intervals may be as short as a few milliseconds and it is a challenge to collect large datasets in time or near-real-time while processing such data.

Table. II: Data class and Data Source of Velocity

SL. No	Data Class	Data Source	Volume
--------	------------	-------------	--------

1	Utility Measurements	Phasor Measurement Unit (PMU)	240 samples/sec
		Smart Meter (SM)	Every 5-15 minutes
2	Weather Data	Radar	Every 4-10 min
		Satellite	Every 1-15 min
		Automated Surface Observing System (ASOS)	Every 1 min
3	Vegetation and Topography	Ecological Mapping System of Texas	Static

iii. Variety: It applies to the numerous data which should be taken into account. Energy networks provide structured data, semi-structured data and unstructured data. It is a challenge to process these different datasets, especially because unstructured data sources are growing. This feature addresses the need for data uncertainty to be addressed. Trust in the veracity of the used data is crucial in a core society market, such as oil, and big data techniques must ensure that truth is one of the most important aspects.

Table. III: Data class and Data Source of Variety

SL. No	Data Class	Data Source
1	Utility Measurements	Phasor Measurement Unit (PMU)
		Smart Meter (SM)
		Intelligent Condition Monitor (ICM)
		Digital Fault Recorder (DFR)
2	Weather Data	Radar
		Satellite
		Automated Surface Observing System (ASOS)
3	Vegetation and Topography	Ecological Mapping System of Texas
		Light Detection and Ranging

iv. Veracity: This function refers to the need to take account of data instability. A new way to collect and analyse data has been introduced by Big Data. For safe operation and stability of the network reliable data and quality statistics on the intelligent grid power network.

Table. IV: Data class and Data Source of Veracity

SL. No	Data Class	Data Source	Veracity (Accuracy)
1	Utility Measurements	Phasor Measurement Unit (PMU)	Error < 1%

		Smart Meter (SM)	Error < 2.5 %
		Intelligent Condition Monitor (ICM)	Error < 1%
		Digital Fault Recorder (DFR)	Error < 0.2 %
		Radar	1.2 dB ms-1
2	Weather Data	Satellite	VIS<2% ; IR <1-2K
		Weather Forecast Model (WFM)	Varies by parameter

These four features enable the use of Big Data techniques in order to extract better performance. Therefore value is frequently designated as fifth V, taking account of the specific application in the field of the power system, these 5V charters will plays a significant role in the system to bring the value added benefits to the system.

3.3 Important applications of Big Data in Power System Sector.

The power supply grid is an extensive electrical network used for the generation of electric power, transmission, distribution and utilization. This sections shows how broad big data analysis in the power system can be integrated with a well-structured way to the various potential applications into the power system. Respects the total of the data set, not its parts. Learning the various divisions of the distribution dynamic data set leads to separate applications that means supporting applications for large data planning. The possible role of power system wide data analytics [20-24] in the following area of activity;

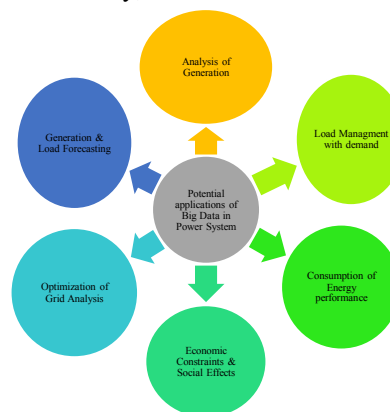


Fig. 6. Important application of Big Data in Power System Sector

4 Big Data Analytics & Machine Learning for Power System.

4.1. Analytics Techniques for Power System.

Big data analysis technology is an important element of sustainable power & energy systems development [6, 25] as described below in the main technologies are;

- Integration and control of electric power in large data networks: The integrated data and

management systems for electrical power in Big Data consisting of the data integration and incorporation. Multiple data sources are now integrated to provide information that is more reliable, accurate and useful than any other data source [26].

- b. Technical analysis of the electricity network data: The method of cleaning, processing and data analysis to analyse valuable knowledge for system decision taking. The core objective of this technique is to extraction of useful data from the data and the data analysis, decision making needed. A large data analysis approach includes patterns analysis, computer training and data mining techniques. The application of data analysis [25, 27] helps to obtain valuable knowledge from mass data.
- c. Large data processing technology for electrical power: Big data technology is important for the processing of data in electricity systems in the electric power network. Data processing is very important for the system, power companies limit their access to large amounts of information that can impact their operation, maintenance and even critical analysis. The processing of data is the processing and understanding of all data. The treatment of data includes computer distribution, memories and stream processing
- d. Visualization of obtained data in power system: Due to the growing value of the information, data visualization is a critical component in the arena of power systems. The graphical representation of data and information is the visualisation of the data. Use different visual devices, such as chart, charts, tables, maps, infographics and dashboards. Tools to view and understand charging, electrical generation, fault prediction offer an accessible way to see and understand trends [28-29].

4.2. Machine for Power System application.

Once the huge data is collected and undergone with various stages and processed and identified the potential data, we need to go for the decision making with the help of advance computing techniques to for the proper valid decision for policy making, marketing, system and device control for the better and operation and maintenance of power system. To this we are having wide variety of Data analytical tools are available which are open source/ as well paid software's such as Apache Spark, Hadoop, Cassandra, Storm, Rapid Miner, R Computing Tool, Neo4j and many more. The most widely used algorithms for data extraction / machine learning generally are referred to as supervised or unsupervised learning, depending on whether each dimension is included. The machine learning algorithms are classified as shown in the below figure 7.

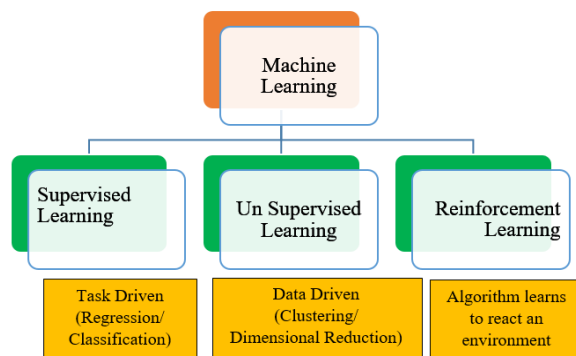


Fig. 7. Classification of machine learning algorithm

In supervised learning the machine is taught by demonstration. They learn from historical data and apply the information of current data in order of forecast future events. Input as well as desired data can assist in the prediction of future events in this scenario. The input of the data is marked or labelled as the correct response. The Unsupervised learning is a way of using data that is not classified or computer-identified. This guarantees that no training data can be generated and that the machine is qualified. The computer must be aware of the data classification without any knowledge about the data. The goal is to incorporate large quantities of data into the machines and to improve their awareness by making previously unknown observations and identifying secret designs. Therefore the unregulated effects Algorithms for learning are not well described. Rather, it is decided what is special or important from the data set [15], [28-30]

In reinforced learning is a sort of machine-learning technique that helps a manager learn from input tests and mistakes, his / her activities and his / her experience in an immersive environment. In comparison to controls where learners have the same number of steps to accomplish a task in the controlled area of learning both supervised and enhanced learning use input and output mapping, enhanced learning uses rewards and penalties to show positive and bad behaviour.

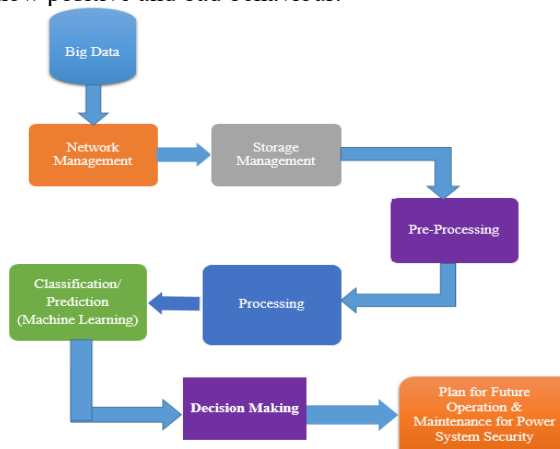


Fig. 7. Complete process flow of diagram big data integration in power system operation

The above fig 7. Shows clearly the process flow diagram of complete path of Big Data Analytics can be incorporated to the power system application.

Conclusion

This article offers an insight into Big Data Analytic roles in the power system, as well as data from different sources at different stages to gather and process these data effectively. It only shows that if we collect information that is not used, it must be processed with standard computing techniques and used to make the right decisions for machine learning such as policy making, marketing, system and devices for better functions and maintenance of the power system. The technology Big Data Analytics can boost knowledge and response times for device operators and can enable the electricity systems industry to identify and respond proactively, use various data analysis techniques, to achieve the reliable operation and maintenance of a power system at different levels.

Acknowledgement

The authors would like to thank HOD, Dean and Management of Presidency University, Bengaluru for giving them the opportunity to work on these particular aspects. Authors also recognize with sincerity all the authors mentioned in this paper's reference section for their worthwhile and sincere efforts to produce such useful articles that are of great interest to young people about the Big Data application in Power & Energy system.

References

1. https://en.wikipedia.org/wiki/Electricity_sector_in_India#Electricity_transmission_and_distribution
2. Mladen Kezunovic, Le Xie and Santiago Grijalva, "The Role of Big Data in Improving Power System Operation and Protection", 2013 IREP Symposium-Bulk Power System Dynamics and Control –IX (IREP), August 25-30, (2013), Rethymnon, Greece.
3. <https://www.simplilearn.com/big-data-applications-in-industries-article>
4. Yingjie Yan, Gehao Sheng, Robert Caiming Qiu, Xiuchen Jiang, "Big Data Modeling and Analysis for Power Transmission Equipment: A Novel Random Matrix Theoretical Approach", DOI 10.1109/ACCESS. (2017).2784841, IEEE Access.
5. S. S. Refaat, A. Mohamed and H. Abu-Rub, "Big data impact on stability and reliability improvement of smart grid," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, (2017), pp. 1975-1982, doi: 10.1109/BigData.2017.8258143.
6. Yang Zhang, Tao Huang* and Ettore Francesco Bompard, "Big data analytics in smart grids: a review" Energy Informatics (2018) 1:8 <https://doi.org/10.1186/s42162-018-0007-5>
7. Lichen Zhang , "Designing Big Data Driven Cyber Physical Systems Based on AADL" (2014) IEEE International Conference on Systems, Man, and Cybernetics October 5-8, 2014, San Diego, CA, USA.
8. Shady S. Refaat, Haithan Abu-Rub, Amira Mohamed, "Big Data, Better Energy Management and Control Decisions for Distribution Systems in Smart Grid ", (2016) IEEE International Conference on Big Data (Big Data).
9. Bo Yang, June Yamazaki, Nao Saito, Yutaka Kokai, Da Xie, "Big Data Analytic Empowered Grid Applications - Is PMU a big data issue?", 978-1-4673-6692-2/15/\$31.00 (2015) IEEE.
10. Vladimir Hahanov , Wajeb Gharibi , Eugenia Litvinova , Svetlana Chumachenko "Big Data Driven Cyber Analytic System", (2015) IEEE International Congress on Big Data
11. Ni Bin, "Research on Methods and Techniques for IoT Big Data Cluster Analysis", (2018) International Conference on Information Systems and Computer Aided Education (ICISCAE).
12. Jun Zhu, Eric Zhuang, Jian Fu, John Baranowski, Andrew Ford , James Shen , "A Framework-Based Approach to Utility Big Data Analytics", IEEE Transactions on Power Systems, Vol.31, No.3, May (2016).
13. Nazreen Junaidi, Mohamed Shaaban, "Big Data Applications in Electric Energy Systems", 978-1-5090-6664-3/17/\$31.00 (2017) IEEE.
14. Olivera Kotevska, A. Gilad Kusne, Daniel V. Samarov, Ahmed Lbath, and Abdella Battou, "Dynamic Network Model for Smart City Data-Loss Resilience Case Study: City-to-City Network for Crime Analytics", IEEE Access Digital Object Identifier 10.1109/ACCESS. (2017).2757841
15. Yuanjun Guo , Zhile Yang , Shengzhong Feng, and Jinxing Hu, "Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms: A Review and a Case Study",
16. Jason Ausmus, Ricardo Siqueira de Carvalho, Aoxia (Kevin) Chen, Yaswanth Nag Velaga, Yingchen Zhang, "Big Data Analytics and the Electric Utility Industry", 978-1-7281-1607-5/19/\$31.00 (2019) IEEE.
17. Arghandeh, Reza & Zhou, Yuxun. (2017). Big Data Application in Power Systems.
18. <https://www.nist.gov/engineering-laboratory/smart-grid/smart-grid-beginners-guide>.
19. <https://sites.google.com/a/aicte-india.org /swayam-repo/big-data-analytics-for-smart-grid>.
20. Xuesong Zhou, Kang Li, Youjie Ma and Zhiqiang Gao, "Research Review on Big Data of the Smart Grid" Proceedings of (2018) IEEE International Conference on Mechatronics and Automation August 5 - 8, Changchun, China.
21. Arbër Perçukua, Daniela Minkovskab and Lyudmila Stoyanovac "Big Data And Time Series Use In Short Term Load Forecasting In Power Transmission System", The 9th International Conference on Emerging Ubiquitous Systems and

- Pervasive Networks EUSPN (2018), *Procedia Computer Science* 141 (2018) 167–174.
22. Gerard Mor, Jordi Vilaplana, Stoyan Danov, Jordi Cipriano, Francesc Solsona, Daniel Chemisana. “EMPOWERING, a smart Big Data framework for sustainable electricity suppliers” (2018), DOI 10.1109/ACCESS.2018.2881413, IEEE Access.
 23. Yi Wang , Qixin Chen , Tao Hong , and Chongqing Kang , “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges”, *IEEE Transactions on Smart Grid*, Vol. 10, No. 3, May (2019).
 24. M. Kezunovic, L. Xie and S. Grijalva, "The role of big data in improving power system operation and protection," (2013) IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid, Rethymno, 2013, pp. 1-9, doi: 10.1109/IREP.2013.6629368.
 25. Jie Zhan, Jinxin Huang, Lin Niu, Xiaosheng Peng, Diyan Deng, Shijie Cheng. “Study of the key technologies of electric power big data and its application prospects in smart grid” 978-1-4799-7537-2/14/\$31.00 (2014) IEEE.
 26. Benjamín Zayas, Alfredo Espinosa, Vicky Sanchez and Javier Perez, “Getting ready for data analytics of electric power distribution systems”, *International Journal of Computers*, ISSN: 2367-8895, Volume 2, (2017).
 27. X. Zhou, K. Li, Y. Ma and Z. Gao, "Research Review on Big Data of the Smart Grid," 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, (2018), pp. 2030-2035.
 28. Yan Sun and T. J. Overbye, "Visualizations for power system contingency analysis data," in *IEEE Transactions on Power Systems*, vol. 19, no. 4, Nov. (2004), pp. 1859-1866.
 29. J. Zhu, E. Zhuang, J. Fu, J. Baranowski, A. Ford and J. Shen, "A Framework-Based Approach to Utility Big Data Analytics," in *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2455-2462, May (2016).
 30. S. A. Shafaat, P. V. Teja, P. S. K. Reddy, R. Manoj, S. B. Daram and O. Obulesu, "Application of Big Data Analytics in Power System under Single Transmission Line Outage Condition," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, (2019), pp. 377-382.