

Application of Big Data in Smart Grids: Energy Analytics

Azamat Marlen*, Askar Maxim*, Ikechi A. Ukaegbu*, H. S. V. S. Kumar Nunna*

*Department of Electrical and Computer Engineering

Nazarbayev University, Astana, Kazakhstan

azamat.marlen@nu.edu.kz, askar.maxim@nu.edu.kz, ikechi.ukaegbu@nu.edu.kz, sivanand.nunna@nu.edu.kz

Abstract—The data that can be extracted from Smart Grids contains a lot of valuable information that would expose hidden opportunities to efficient utilization of existing resources. The data comes in the form of measurements from smart meters/devices. Energy Analytics is irreplaceable for this kind of task. It is a process of collecting data from smart meters/devices both real time and from archival stores and applying some sort of data analysis technique to gain some insight into important correlations, trends, and patterns. This paper offers elaborate discussion on the application of big data in smart grids and describes future prospects for this technology in Kazakhstan. Furthermore, this paper demonstrates a case study of load profiling using the data set that contains energy consumption readings for London households between November 2011 and February 2014. To analyse the data set the K-means clustering algorithm for unsupervised learning is used here. The results of this overview clearly demonstrated the fields where energy analytics can impact the energy segment of Kazakhstan infrastructure and introduces possible ways of overcoming challenges that are present in power systems. However, it is apparent that, in order to realize all these possibilities, it is important to increase the awareness of both government (i.e. the ones to implement the policy) and citizens (i.e. the ones who would make it possible).

Keywords— Big Data, Energy Analytics, Kazakhstan, Load Profiling, Smart Grid

I. INTRODUCTION

In current reality, where the demographical and economical growth calls for an increased demand of electricity, and newly developing industries require efficient energy management systems, conventional electric power infrastructures are struggling to provide sufficient service. As one of the most important systems, which are responsible for delivering the electricity over the high voltage transmission lines from remote power plants to individual customers, electrical grids are no exception. The presence of environmental and economical concerns, the emergence of alternative power generation technologies, electric vehicles, and other technological innovations have made these unidirectional and centralized systems outdated and not well suited to meet the reliability, operational efficiency and sustainability requirements posed by those changes, which in turn prompted the need for developing new means of energy networking [1], [2].

With the inception of the Smart Grid era and increased availability of equipment that produces renewable energy (e.g., solar panels, wind turbines), households and small businesses that were usually regarded as passive consumers, were able to actively contribute the excessive electricity that they generated for the consumption of other people [3]. The basis of this technology is bidirectional communication and control between producers and consumers, such that both power and data can be easily transferred. This new feature implies two things: first, there would be a huge amount of data that smart grid produces

during the course of its work; and second, that Big Data concept can find its application in smart grids. That's where the idea of Energy Analytics is introduced. It is a process of collecting sensor data both real time and from archival stores and applying some sort of data analysis technique to gain some insight into important correlations, trends, and patterns.

This paper offers elaborate discussion on the application of big data in smart grids, and describes future prospects for this technology in Kazakhstan. The rest of the paper organized as follows. Section II presents the background information and introduces all the important concepts and ideas. Section III describes the stages of energy analytics. Section IV discusses possible applications of energy analytics. Section V describes challenges that are present in the field. Section VI discusses future prospects of this technology in Kazakhstan. Finally, section VII concludes the paper.

II. BACKGROUND

A. Big Data

In the present time, the term of big data is on everyone's lips. However, no one knows where the concept was started. The first mention of big data was made by Alvin Toffler in his written work "The Third Wave" in 1980s [4]. Further, during the next decade, many Chinese researchers discussed the application of a large amount of data and its benefits to society [4]. Obtaining of such quantity of data was unrealistic due to insignificant development in technological field. However, with the progress of computational equipment and devices such as Internet, mobile phones and tablets, the commercial and business spheres started to significantly highlight the value of digitization and informatization in the industry. The result of these actions was a huge quantity of generated data or, in other words, Big data.

The value of big data is difficult to understand. The point is that Big data does not only refer to the quantity of data. It has diverse features such as value, volume, variety, velocity and veracity [5]. The collected data is becoming more beneficial for industrial companies for decisions that lead to total revenue growth. Information about business operation is subjected to further analysis of past failures, modeling future operations and probability prediction of repetition of such events. For example, in 2012, USA government spent more than \$200 million to research analyzing techniques of Big data to improve the quality of security and accelerate the speed of innovations in related spheres [5].

B. Smart Grid

For many years, conditional grids were the main source of power. However, current progress leads to significant changes. A prime example is Smart grid. Smart grid can be defined as a new generation of grids that consists of hardware, software technologies and communication network [6]. Current high demand in power sources

leads to overload of conventional power grids. As a result, system's failure is happening frequently. These circumstances affect stability of power supply and economic situation in certain regions [7]. While, the main purpose of Smart grid is to provide high reliability of system and flexibility. Also, the smart grid is able to reduce power losses happened at transmission and distribution processes. Moreover, in the last decade, the point of renewable energy becomes more attracting day by day. The novel feature of new grids is that renewable sources like windmills and photovoltaic plates can be easily connected to a distribution system to decrease overload at power grids [6]. In other words, the smart grid gives opportunity to develop renewable sources of energy.

The progress that can be observed in smart grid systems is the results of development in power industry that took place during the last several decades. Not only processes of generation, transmission, distribution and consumption were revised and redesigned to perform higher efficiency, but, also, sensor and metering devices were implemented to supervise the operation of grid.

Smart grid promotes the development of renewable energy sources. This led to the emergence of a new concept of distributed power generation [6]. The influence of renewable power sources such as wind, sun, a biomass waste, and geothermal on a power system is becoming more significant. However, such energy sources might be unreliable due to different reasons. For example, on cloudy day, it is impossible to obtain energy from photovoltaic plates. The merit of Smart grid is that system is able to deal with problems of distributed generation and increase the quality of distributed energy. Moreover, in smart grid systems operations related to renewable energy sources are monitored by sensors and smart meters.

C. Big Data & Smart Grid

To characterize the concept of Big data, the conventional model of 5V's is used worldwide by many scientists in different areas [8], [9]:

- **Volume** is responsible for an amount of generated samples. In smart grids, the measurement sources of data such as sensors and smart-meters easily provide enormous number of data samples. The storing of such huge data is becoming a substantial problem. For this situation, different storage places are provided. Furthermore, these locations are connected together through software so that the data in one storage can be easily accessed by another one.
- **Velocity** is responsible for speed at which data is generated and transferred further. The problem is that every sample has its own weight. For instance, in [8], 1 million smart-meters were programmed to send 4 readings per hour which resulted in 2920 Tb of data at the end of the year. Storage of such amount of data and data transferring seems to be really problematic in the future.
- **Variety** indicates wide choice of data sources. In present time, various sources of data such as sensor readings, videos or messages in social media might be processed to obtain more reliable results than it was before.
- **Veracity** refers to the quality of obtained data. In such huge amount of gained data, the probability that this information may influence the final result of data analysis is huge. The inaccurate measuring might be the result of sensor failure or transferring problems.
- **Value** is responsible for subtracting beneficial data from given data set. The next step is to understand the importance of data and how it might be valuable. The issue is that, as the data set gets bigger, the amount of important information decreases. However, by the implementation of more accurate intelligent metering devices, the amount of valuable data becomes larger.

D. Data sources in Smart Grids

The smart grid is an intelligent system that subtracts the enormous amount of data from operational processes of system. To be more precise, data is collecting from generation of power until the consumption by customers. Moreover, such data contains readings from smart meters, economic situation, and weather condition of a certain region [10]. As well as in other fields, big data provides opportunity for choosing an appropriate location for power grid, a proper use of power equipment, and increasing revenue by analyzing gained data.

The data collected in smart grids might be separated into three groups: measurement, business, and external data [6]. The huge part of data collection operation is performed by different kind of sensors that are installed in the system and smart meters in customer's homes, offices or factories. This is called a measurement source of data. The business data source contains information about economic strategy and its failure experience. Information about weather or holidays cannot be collected, but such external source of data affects the operation of power grid [11].

III. ENERGY ANALYTICS

A. Data collection, communication, & preprocessing

As it was mentioned in the previous part, depending on the source of the data, means of data collection vary. For instance, the data pertaining the elements outside of the smart grid (e.g., weather, electricity prices, etc.) are usually collected from various outside sources, whilst energy related quantitative data in smart grids is collected and transmitted with the help of smart meters, that are installed in both the utility companies and customers. In these systems, depending on the design, measurements are taken with a different level of accuracy and sampling resolution. Typically, these measurements include details such as the feeder current, energy over a period, node voltage, power factor, total harmonic distortion, active and reactive power, load demand, etc. [11].

After the data has been collected, it is essential to ensure stable and cost-effective communication infrastructures, that would be able to connect sources of the data, that are distributed over large territories. These infrastructures can be both wired and wireless, and each of them has their own merits and demerits. With wireless communication the data from numerous smart meters can be transmitted to data center at low cost, even though the data might be polluted with noise. On the other hand, wired communication technology, such as Power Line Communication (PLC) is already successfully implemented [11].

There are 3 types of networks that comprise the communication infrastructure for smart grids [12]:

- **Home Area Networks (HAN)**, is a network of smart meters and devices within a smart house or office. With a data transmission rate of less than 1 Kbps, it is usually applied for local energy management.
- **Neighborhood Area Networks (NAN)**, is a network of several HANs, whose main function is collection and storage of energy consumption data in Load Data Center (LDC).
- **Wide Area Networks (WAN)**, is a network of all smart grid components. Covers a large area and has large data transmission capability.

By this stage it became apparent that acquiring and processing all the data that smart grid generates is not the wisest choice, and doing so is not going to affect anything, other than increasing a cost of communication, computational complexity, and data storage resource utilization [13]. To avoid such unfavorable outcomes and increase the speed and efficacy of data analytics, the data preprocessing is required. Data preprocessing is a process when the dimensionality of the data is reduced so that the further classification of the data will have lower true error, or to achieve more compact, easily understandable representation of the target concept [14]. In [13] data preprocessing techniques were generalized into 3 categories:

- **Data Integration**, which is about identifying similar features that has different names, or different features with similar names, which came as a result of combining all smart meter data sets into one, and reducing the data set by abandoning those "extra" features. It also deals with features that are highly correlated.
- **Data Cleansing**, which is about dealing with missing and abnormal values. In case of missing values, most frequently deletion (i.e. getting rid of the sample point entirely) or interpolation (i.e. generate a new feature value using values from other samples) techniques are used. Abnormal values, on the other hand, are either treated as a missing value or kept in the data set considering that it was a very rare event that just happened.
- **Data Transformation**, which is about feature extraction, where a subset of new features is created by combinations and transformation of existing features.

These techniques are presented in generalized form, and depending on the requirements pertaining to certain tasks and the nature of data set they can be used independently or in combination with one another.

B. Data analytics

Data analytics or machine learning is about extracting valuable information from the sea of organized data for the purpose of discovering patterns, trends, and correlations, that would expose hidden opportunities to efficient utilization of existing resources. Within the context of smart grids, comprehensive data analytics algorithm is what makes smart grids smart, and without it achieving clean, reliable, and efficient grid is a far-fetched dream [15].

In a machine learning the most common approaches are supervised and unsupervised learning. The difference between the two is that, in supervised learning the focus is on class affiliation, while in unsupervised learning the focus is on innate structures of the data. In other words, in supervised learning the data is labeled, and in unsupervised learning it is not, so it is necessary to look into data separability to gain any insight from it [16]. Some of the popular algorithms for supervised and unsupervised learning are [11]:

- Decision Tree
- Naive Bayes
- Support vector machine classifier
- K-Nearest Neighbor (KNN)
- Random Forest
- K-means
- K-medoids
- Hierarchical Clustering
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- Expectation-Maximization

Similar to the data preprocessing, suitable algorithm selection is closely tied to the requirements pertaining to certain tasks and the nature of data set.

IV. APPLICATIONS

Let's take a look at the topic of energy analytics' application in smart grid:

- **Fault Detection**
Conventional methods of fault detection in grids are based on identifying location of overcurrent in the system. For Smart grids, [17] suggests that wavelet transform and decision tree might be used for detecting fault. Besides, in several protection schemes, data should be taken from both sides of a certain power supplier. Furthermore, discrete Fourier transform is applied to data set and based on a decision tree model, a final decision

on relating condition is made [18]. There are various way of detecting faulty conditions in the system based on energy analytics and all of them are widely used in smart grids to avoid failure of system.

- **Power Quality Monitoring**

The topic of power quality is of a high significance in power systems and it has huge impact on the safe operation of devices. This is further accentuated by the fact that, harmonics distortion problems might be faced because non-linear devices are becoming widespread. To eliminate problems, these deep learning techniques are suggested. One of the most successful is a technique that works on the image of signal's waveform [19]. Merit of the proposed method is in the fact that, there is no necessity to record data samples of current and voltage. The method is performing with accuracy of 100%.

- **Load Forecasting**

Load forecasting is considered as crucial feature of smart grids that greatly affects energy management. The deep learning method is accounted as highly efficient way for long-term load forecasting [20]. Successful and accurate results of consumption prediction may affect to electricity market positively and financial merits of energy consumption forecasting must be mentioned [4]. Furthermore, it should be mentioned that weather condition has significant effect to the load forecasting. For example, typically, power demand in winter is considerably higher than in summer [10].

- **Load Disaggregation**

With the help of energy analytics, it is possible to obtain data on electricity consumption of certain device. This feature is called load disaggregation. One of the most widespread techniques is related to harmonics detection of current in the system [21]. Also, different ways of load disaggregation might be based on machine learning algorithms or fuzzy logic.

- **Non-technical Loss Detection**

There are two types of non-technical loss: theft of electricity and bill calculation errors. The financial losses is considerably huge. For example, in 2014, electricity losses reached more than \$85 billion worldwide [11]. However, based on energy analytics, approaches on support vector machine and decision tree were suggested by [22]. Firstly, collected data is processed by decision tree to obtain approximate energy consumption quantity. Further step is to proceed this data to support vector machine where final decision is made.

- **Load Profiling**

Distribution and retail companies that are able to segment the market and target those segments would be able to develop better marketing strategies and define specific contracts based on different consumption patterns. Load profiling enables this segmentation by clearly demonstrating different patterns of electricity demand by different groups of customers. These patterns are derived from daily load diagrams. For the efficient implementation of this application, involvement of energy analytics is a huge boon.

Energy analytics have more application in a sphere of power system. Besides mentioned applications, there are health monitoring and others [11].

V. CHALLENGES AND POSSIBLE SOLUTIONS

Big data is an essential finding of present time. It has many applications in various sphere of industry. In this paper, the big data application in smart grid is discussed. Despite the fact that big data analytics has significant advantages, several challenges might be encountered during the operation of big data sets.

- 1) One of real problems of data sets is that data is coming from diverse types of sensors and actuators in smart grid.

To deal with a specific set of data that has various forms, data integration is a valid option. Another problem is that significantly huge data storage is required for it.

- 2) In some cases of smart grid operation, an immediate response of the system is necessary. Example of it might be a fault condition in electrical power system. To detect extraordinary behavior of the system, real-time processing is required. It has several beneficial sides such as swift data processing, a particular system's algorithm to deal with big data sets to avoid fault in operation of smart grids.
- 3) Another challenge of big data analytics is that any slight alteration in the system's operation must be shown graphically. The data is coming from different destinations and it makes this problem notable. The software equipment with high accuracy has to be implemented.
- 4) Data privacy is considered as one of the considerable challenges in big data analytics. Successful cyber attacks might lead to violation of private security, huge bills and many other cascading problems.

There are ways to tackle challenges of big data in energy analytics. The issues of data variety and volume might be solved by several software programs. The most reliable among them is Apache Hadoop software platform. The main operation purposes of software are to work with massive data and separate it among infrastructures for its processing within short time period. The software operation is simple and can be studied. For the problem of velocity, it is possible to offer the approach based on a hybrid of cloud computation and SaaS/PaaS/IaaS models. Besides the problem of velocity, this method can be beneficial for data security and privacy. Also, the problem of privacy might be solved by encryption. For instance, one of the encryption methods such as Attribute-Based Encryption might be used to obtain more protective access to encoded data. In current circumstances, proposed solutions for challenges are considered as the most effective and easily accepted ways. However, further investigation in the topic of Big data challenges might result in more efficient and effective approaches.

VI. LOAD PROFILING

A. Database

To perform a case study, a specific dataset that satisfies the requirements of the task is needed. For this purpose "Smart-Meter Energy Consumption Data in London Households" was used. The data set contains energy consumption readings for a sample of 5567 London households between November 2011 and February 2014. Readings from smart meters were taken at half hourly intervals. The dataset contains following information:

- **LCLid** is an identification number of a household.
- **stdorToU** tells us what kind of tariff was applied. There are two of them: "ToU", which is dynamic tariff; and "Std", which is just a classic fixed tariff).
- **DateTime** is date and time of sampling.
- **KWH/hh (per half hour)** is an energy that was consumed in the past 30 minutes in kWh.
- **Acorn & Acorn grouped** is an "Acorn" group associated, that categorized households.

For our purpose, only "DateTime" and "KWH/hh (per half hour)" are required. To remove all the unnecessary details, the dataset is to be reduced. A python code that removes these details from .csv file is implemented. At the end we get dataset where only date and energy consumption in kWh remained. Furthermore, the data from 10 households with each having 17457 samples are used.

B. Results

The load profiling is performed as a case study of big data application. To analyze data set for household energy consumption,

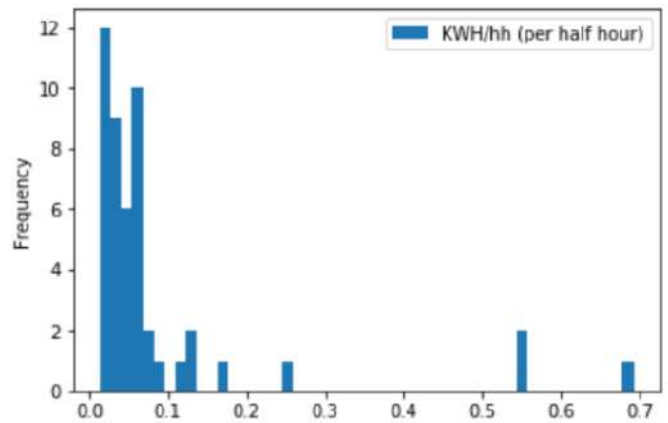


Figure 1: Low consumption household

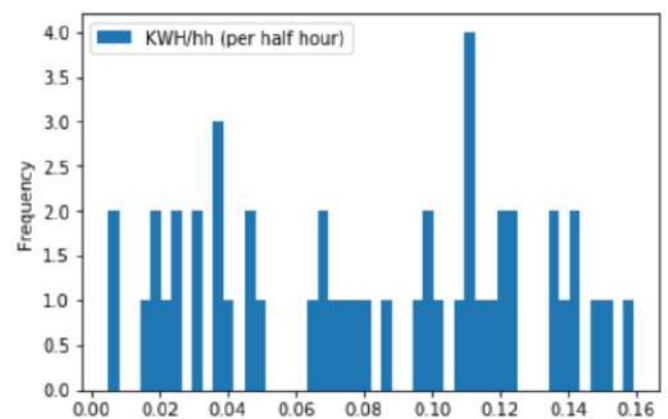


Figure 2: Medium consumption household

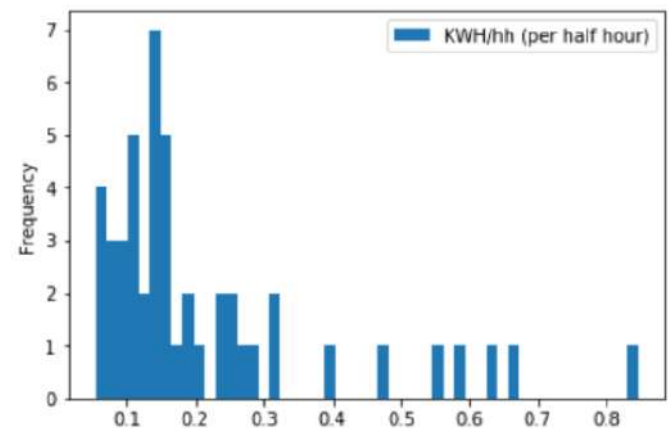


Figure 3: High consumption household

K-means clustering algorithm for unsupervised learning is utilized. Clustering Dispersion Indicator showed high values for low number of clusters. Elbow method is used to determine optimal number of clusters for accurate analysis. The K-means method is implemented on 6 month of winter and summer seasons of 2013 to observe the difference between consumption patterns. The following observations

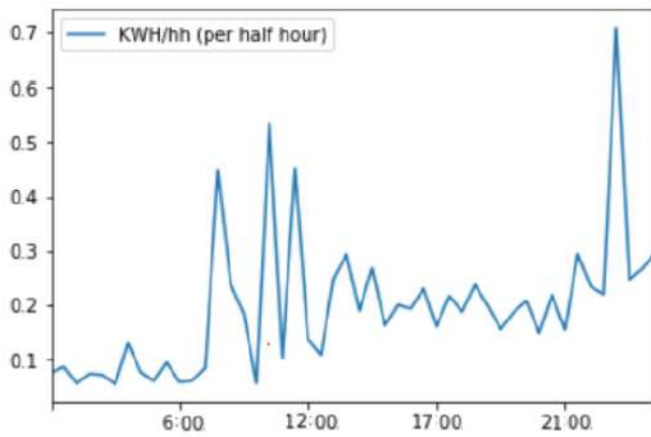


Figure 4: Summer daily load diagram of the household

were made:

- Energy analytics showed great efficiency at determining the trends for energy consumption among households. There were clear distinctions between low consumption (fig. 1), medium consumption (fig. 2), and high consumption households (fig. 3).
- Clear observations on the dependence of energy consumption on the season were made. For instance, figure 4 shows the daily load diagram of the household during the summer. It can be noticed, that throughout the day consumption is quite uniform and there are two periods when consumption is at the peak: morning and late evening.

One of the main benefits of load profiling is that it can be applied to demand response. Load profiling helps to power system avoid overload condition at peak hour that usually in the morning and evening. There are two ways deal with peak hour energy consumption issue such as price-based and incentive-based [23]. The price-based approach can be described as increasing power prices at peak hour. Whereas, incentive-based method is working remotely by switching on or off consumer's utilities to decrease amount of energy consumption.

VII. FUTURE PROSPECT OF ENERGY ANALYTICS IN KAZAKHSTAN

The applications that has been discussed in section IV can play an essential role in the development of Kazakhstan's electricity market. There are two factors to this statement:

- Increased energy consumption
- Renewable energy sources

This increase in energy consumption stems from the fact that the main contributor to the rapid development of country's economy is industry, in particularly the extraction of natural resources (e.g., oil, coal, etc.) consumes a large amount of electricity. Between years 2005 and 2017, within the span of 12 years, energy consumption in Kazakhstan increased by 44.54% [23], [24]. in 2010 alone, 85.3% of the total energy consumption was contributed by the industry [23]. These statistics tells us that further development would increase demand for energy. And whilst further increasing the means of energy production would require large investments, improving energy management by means of energy analytics is more plausible in a short term. In addition to that, it can also contribute to the establishment of firm power reserves, and consequent self-sufficient power systems.

Astana EXPO 2017, which as the name suggests took place in the capital of Kazakhstan, Astana, and was dedicated to the topic of

future energy, which is largely about renewable energy sources. This event signified that the government is serious about supporting the development of this industry in Kazakhstan. This intention is further confirmed by the fact that in 2017 1.109 Billion kWh of power was generated from renewable energy sources [24]. It made up 1.1% of total energy that was produced during that year, compared to the less than 0.5 % that was in 2011, there is some progress [24], [23]. However, the issue is that the majority of energy production is still attributed to natural gas and coal burning, and this in turn gave rise to a major environmental concern. Implementation of energy analytics might increase the efficiency of renewable energy production.

As it can be seen from presented case study, load profiling can clearly show peak electricity demand period of time. By this data, it is possible to manage power facilities and decrease overloading of power grids and save investment money. It is one of the possible implementation examples of Energy Analytics field. Further investigation in the topic in Kazakhstan may lead to significant saving and increasing of efficiency of supplied power. More research should be done in the power field in Kazakhstan to detect the possible ways of Big data gathering implementation to optimize energy industry industry field. The example of Kazakhstan might be first step of Energy Analytics application in the Central Asia.

VIII. CONCLUSION

In this paper, detailed introduction to the energy analytics process was presented, and techniques that are used at each stage was discussed. Case study of Energy Analytics was performed as an example of Big data application in power industry. Furthermore, this paper describes future prospects for this technology in Kazakhstan. The results of this overview clearly demonstrated the fields where energy analytics can impact the energy segment of Kazakhstan infrastructure and introduces possible ways of overcoming challenges that are present in power systems. However, it is apparent that, in order to realize all these possibilities it is important to increase the awareness of both government (i.e. the ones to implement the policy) and citizens (i.e. the ones who would make it possible).

REFERENCES

- [1] E. Ancillotti, R. Bruno, and M. Conti, "The role of communication systems in smart grids: Architectures, technical solutions and research challenges," *Computer Communications*, vol. 36, no. 17-18, pp. 1665-1697, Nov. 2013.
- [2] M. Rajarshi, N. Sarda, and H. Kargupta, "Distributed data mining for sustainable smart grids," *Proc. of ACM SustKDD* 11, pp. 1-6, Aug. 2011.
- [3] P. Goncalves Da Silva, P. G. Da Silva, D. Ilic, and S. Karnouskos, "The Impact of Smart Grid Prosumer Grouping on Forecasting Accuracy and Its Benefits for Local Electricity Market Trading," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 402-410, 2014.
- [4] H. Zhang, Y. Li, C. Shen, H. Sun, and Y. Yang, "The Application of Data Mining In Finance Industry Based on Big Data Background," in *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*, pp. 1536-1539, 2015.
- [5] L. Yang and J.-J. Zhang, "Realistic plight of enterprise decision-making management under big data background and coping strategies," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp. 402-405, 2017.
- [6] S. Guner and A. Ozdemir, "Turkish Power System: From conventional past to smart future," in *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, pp. 1-4, 2011.
- [7] A. Zahedi, "Developing a system model for future smart grid," in *2011 IEEE PES Innovative Smart Grid Technologies*, 2011.
- [8] S. Sagiroglu, R. Terzi, Y. Canbay, and I. Colak, "Big data issues in smart grid systems," in *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 1007-1012, 2016.
- [9] T. Zhu, S. Xiao, Q. Zhang, Y. Gu, P. Yi, and Y. Li, "Emergent Technologies in Big Data Sensing: A Survey," in *International Journal of Distributed Sensor Networks*, vol. 11, no. 10, pp. 1-13, 2015.

- [10] B.-J. Chen, B. -J. Chen, M. -W. Chang, and C. -J. Lin, "Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001," in *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821-1830, 2004.
- [11] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in smart grids: a review," in *Energy Informatics*, vol. 1, no. 1, pp. 1-24, 2018.
- [12] D. Baimel, S. Tapuchi, and N. Baimel, "Smart grid communication technologies-overview, research challenges and opportunities," in *2016 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, pp. 116-120, 2016.
- [13] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, "Big Data Analytics for Dynamic Energy Management in Smart Grids," in *Big Data Research*, vol. 2, no. 3, pp. 94-101, 2015.
- [14] Kotsiantis, S. B., Dimitris Kanellopoulos, and P. E. Pintelas, "Data pre-processing for supervised learning," in *International Journal of Computer Science*, vol. 1, no. 2, pp. 111-117, 2006.
- [15] C.L. Stimmel, *Big data analytics strategies for the smart grid*. Auerbach Publications, 2016, p. 31.
- [16] Zhao Z., Liu H, "Spectral feature selection for supervised and unsupervised learning," in *Proceedings of the 24th international conference on Machine learning*, pp. 1151-1157, 2007.
- [17] D. P. Mishra, S. R. Samantaray, and G. Joos, "A Combined Wavelet and Data-Mining Based Intelligent Protection Scheme for Microgrid," *IEEE Transactions Smart Grid*, vol. 7, no. 5, pp. 2295-2304, 2016.
- [18] S. Kar, S. R. Samantaray, and M. Dadash Zadeh, "Data-Mining Model Based Intelligent Differential Microgrid Protection Scheme," *IEEE Systems Journal*, vol. 11, no. 2, pp. 1161-1169, 2017.
- [19] E. Balouji and O. Salor, "Classification of power quality events using deep learning on event images," in *2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, 2017.
- [20] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network," *IEEE Transactions Smart Grid*, pp. 1-1, 2017.
- [21] M. Berges, E. Goldman, H. Scott Matthews, and L. Soibelman, "Learning Systems for Electric Consumption of Buildings," in *Computing in Civil Engineering (2009)*, 2009.
- [22] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005-1016, 2016.
- [23] Green Bridge Partnership Programme, "Electric power industry of Kazakhstan", 2016. Available at: <http://gbpp.org/ru/2016/12/15682>
- [24] Samruk Energy, "Analysis of the electricity market of Kazakhstan", 2018. Available at: <http://eenergy.media/wp-content/uploads/2018/02/analiz-rynka-elektroenergii-kazakhstana-za-2017.pdf>

IX. BIOGRAPHY



Azamat Marlen is graduate of Nazarbayev Intellectual School in Uralsk. Currently, he is pursuing B.S.degree in Electrical and Computer Engineering department with specialization in power field at Nazarbayev University. His current research interests include renewable energy sources, power management and microgrid fields.



Askar Maxim is currently pursuing B.S.degree in electrical and computing engineering at Nazarbayev University. His current research interests include software engineering field, Internet of Things and Machine Learning field of interests.



Ikechi Ukaegbu (S'M) received his B.Sc. in Electrical Engineering, Electromechanics and Electrotechnology at Moscow Power Engineering Institute (Technical University), Moscow, Russia in 2004 and M.Sc. in Electronics and Microelectronics at the same university in 2006. He obtained his Ph.D. at Korea Advanced Institute of Science and Technology (KAIST) in 2012. He has worked as a Post-Doctoral researcher in Electrical Engineering Department at KAIST from 2012 to 2013. He held RD positions at Electronics and Telecommunications Research Institute (ETRI), Korea, from 2008 to 2009 and at Lightron Fiber-Optics Inc., Korea in 2013. He also worked as a senior engineer with the design technology team at Samsung Electronics Co. Ltd, Korea from 2013 to 2016. He co-founded a startup company where is served as the CTO from 2016 to 2017. He joined Nazarbayev University School of Engineering in 2018 and he is currently an assistant professor with Electrical and Computer Engineering Department. His research interests include 5G Technologies, Electronics & Nanophotonics, Signal & Power Integrity, Channel Modeling, and Terahertz Chip-to-chip links. More information can be found at www.idslab.info



H.S.V.S. Kumar Nunna received his bachelors (B.Tech.) degree in Electrical and Electronics Engineering from the Jawaharlal Nehru Technological University Hyderabad (JNTUH), in 2007, Masters degree (M.Tech) in Industrial Power and Automation (formerly know as Computer Controlled Industrial Power) from National Institute of Technology Calicut, India in 2010 and Ph.D. degree in Power and Energy Systems from IIT Bombay, India in December, 2014. From September 2017 Dr. H S V S Kumar Nunna is an Assistant Professor in the school of engineering at the Nazarbayev University (NU), where he teaches and conducts research in Power and Energy systems as a faculty member of the department of Electrical and Electronics Engineering. Before joining NU, he was with the National University of Singapore (NUS) as a Postdoctoral Fellow in the Center for Green Energy Management and Smartgrids from January 2015 to June 2017, and with Indian Institute of Technology Bombay (IIT Bombay), Mumbai from July 2014 to January 2015 as a Research Associate in the department of Energy Science and Engineering.