# Big Data & Analytics for Wind O&M: Opportunities, Trends and Challenges in the Industrial Internet

**Bouchra Bouqata, Ph.D.**
GE Renewable Energy
Digital
bouqata@ge.com

## ABSTRACT

The popularity of clean renewable energy, such as Wind energy, has been moving to the forefront of social awareness, public policy as well as major tech corporations' investments to power 100% of their data centers from renewable energy [1,2, 3,4]. As Wind energy is becoming more economically competitive, wind farm operators must understand and manage the performance analysis of wind farms in order to achieve desired production and revenue goals. However, farm operators are being faced with a deluge of data from multiple sensors connected to the wind turbines complex systems. Big Data & Analytics are already resulting in disruptive innovation across many industry sectors. Given the uncertainty and complexity associated with wind energy systems, there is huge potential for using these techniques to significantly improve performance and reduce the costs of future wind energy systems. There also a paradigm shift with the Internet of Things (IoT), connecting machines to machines through networks, data and analytics, as another important technology to deal with challenges of Big Data Analytics for Wind O&M. IoT will enable big data analytics at large scale to address complex systems such as wind farms Operations and Maintenance (O&M). Through the emerging technologies in IoT's Advanced Analytics capabilities such as GE's Asset Performance Management Diagnostics and Prognostics [5,6,7], it is possible to drive operating expenses savings and move away from traditional reactive O&M to sophisticated predictive and proactive O&M solutions. Somewhat unique in wind energy, this approach extensively leverages physics-based modeling of the system and fuses Physics Based modeling along with Data Driven models and Statistical and Machine Learning techniques. This summary paper will discuss current trends and enabling technologies in Big Data and Analytics in IoT to drive smart decision making. Then, it will describe the opportunities and challenges of this new paradigm to address wind farm Operations and Maintenance

(O&M) expenses and move from reactive to proactive O&M. To conclude, it will outline remaining challenges to be addressed.

## INTRODUCTION

The popularity of clean renewable energy, such as Wind Energy, has been moving to the forefront of social awareness, public policy as well as major tech corporations' investments to power 100% of their data centers from renewable energy [1,2, 3,4]. As Wind Energy has been saturating the market, the more economically competitive it became to increase its output (Performance) and reduce its cost (Maintenance). In order to achieve desired production and revenue goals, wind farm operators must understand and manage the performance analysis of wind farms. However, farm operators are being faced with a deluge of data from multiple sensors connected to the wind turbines complex systems. Big Data & Analytics are already resulting in disruptive innovation across many industry sectors. Wind energy systems stand-out amongst other complex technical systems due to the combination of large levels of wind uncertainty along with the high levels of interaction and coupling of the wind farms physics. This provides a huge potential for using Big Data Analytics techniques to significantly improve performance and reduce the costs of future wind energy systems.

Today, Data is estimated to be created at 2.5 quintillion bytes/day from sensors, social media, images, etc. This includes the growing use of Big Data in wind power assets Operations and Maintenance (O&M) that is estimated to be at 25 Trillion bytes/day. The ubiquitous availability of data has created a paradigm shift from information-poor to information-rich and impacting virtually every area of modern life.

This summary paper will, first, provide an overview of Big Data Analytics trends, challenges and enabling technologies at large and more specifically to Wind Energy O&M. Second, it will describe the Internet of Things (IoT), connecting machines to machines through networks, data and analytics, as another important technology to deal with challenges of Big Data Analytics for Wind O&M. Then, it will describe the opportunities and

challenges of this new paradigm to address Wind O&M expenses and move from reactive to proactive O&M. To conclude, the paper will outline remaining challenges to be addressed.

## Big Data Advanced Analytics: Trends and Challenges

Big Data Automated Analytics for knowledge discovery, especially Machine Learning (ML), has emerged as an overarching area to enable knowledge discovery and help make smart decisions. Big Data & Advanced Analytics have been successfully used to address problems in many industrial domains, resulting in disruptive innovation that can be leveraged to solve wind energy performance and maintenance cost challenges. The design and development of high quality large scale analytics is a complex process, largely involving search in a space of "Big" noisy, structured and/or unstructured data sets. Analytics require also searching in a large pool of diverse models. Additionally, evaluating just a single model, would involve a search across all combinations of structures and parameter values. Moreover, finding the right scalable machine learning approach could require many expert efforts. Nevertheless, the availability of new infrastructures at scale such as cloud platforms has given a new direction to solving the challenges stated above. This created a shifting paradigm that needs to involve automating a significant portion of the currently manual process involved in problem formulation (to select the appropriate machine learning algorithms), data preparation, model selection, model tuning, etc. In addition, leveraging parallel computing environments through cloud computing (such as Hadoop), high-performance computing, and large scale optimizations are deemed important to create, maintain and deploy large scale machine learning on Big Data.

Another important aspect of Big Data Analytics is good quality data. The necessity of a good quality data set is obvious when you are building a predictive model. There are two main challenges when dealing with data quality: 1) The data is unlabeled even when there is a large pool of it 2) The features do not have predictive power. In the first scenario, the data could be annotated by subject matter experts. Since their time is very expensive, the challenge is what data points are the most informative to

focus the precious time of experts on. In this case, there are ML techniques, such as Active Learning [8], which is able to interactively query the subject matter expert to obtain the desired outputs at new data points. to solve an optimization problem to get the highest performance with the predictive model with the smallest training set. This is the case of active learning. The second scenario is a common case when the predictive problem is very complicated and you do not have predictive features. In this case, Feature Engineering [9], can help come up with/design the best or a better representation of the sample data to provide the necessary information to the predictive algorithm.

## Big Data Analytics in the Industrial Internet for Wind Operations & Maintenance:

Another important aspect of technology to deal with challenges of Big Data Analytics is the Internet of Things (IoT). It is about connecting machines to machines and to people through networks, data and analytics. It promises to remake global industry, boost productivity, and launch an entirely new age of prosperity and growth. IoT is shaping modern businesses from manufacturing to marketing. Machines/objects have always issued early warnings, but in an inconsistent and un-actionable way. The advent of networked machines with embedded sensors and advanced analytics tools has changed that reality. Most machines/objects now either have or are in the process of getting multiple sensors and being connected. These sensors create a plethora of data sources that are often neither connected nor integrated, as it is the case with the deluge of data from wind turbine sensors. To unleash the power of data integration and systems-level analytics and optimization in applications such as in wind O&M, it is critical to ensure interoperability between data sources. The ubiquitous availability of this digital information has transformed the world as we know it, creating a paradigm shift from information-poor to information-rich, and impacting virtually every area of modern life. Yet, we are still a long way from maximizing the potential of the Information Revolution and effectively using the ubiquity of digital information to empower the Internet of Things. Moreover, several concerns about privacy and cyber

security implications are raised by both industry and government. The risk of connecting unsecure devises to the Internet should be properly mitigated. Therefore, cybersecurity and physical security solutions need to work together. To accelerate secure data-driven innovation and discovery, new technologies, infrastructure (networking, storage architecture, cloud computing), new platforms and cybersecurity technologies are needed to empower industry to tackle this new shift in “Big data” from machines and objects.

The Industrial Internet—the combination of Big Data analytics with the Internet of Things is producing huge opportunities for companies in all industries, but especially in Renewable Energy. As a recent analysis, [5], points to,” Not all Big Data is created equal.” According to the authors, “data created by industrial equipment such as wind turbines, jet engines and MRI machines … holds more potential business value on a size-adjusted basis than other types of Big Data associated with the social Web, consumer Internet and other sources.” To accelerate data-driven innovation and discovery, new technologies and infrastructure are needed to empower industry. For that end, GE has invested significantly in a new Industrial Internet platform, Predix Asset Performance Management [5,6,7], to enable big data analytics at large scale to address complex systems such as wind farms Operations and Maintenance (O&M). Through Asset Performance Management’s Diagnostics and Prognostics capabilities, it is possible to drive operating expenses savings and move away from traditional reactive O&M to sophisticated predictive and proactive O&M solutions. Somewhat unique in Wind Energy, this approach extensively leverages physics-based modeling of the system and fuses Physics Based modeling along with Data Driven models and Statistical and Machine Learning techniques. This provides three important aspects to increase Performance and reduce Maintenance cost in Wind Energy O&M:

- Continuously collects data from assets combined with other operational data to monitor, analyze and improve performance and maintenance

- Delivers insights through applying asset0-specific advanced analytics models
- Provides in advance the asset issues to make smart decision and execute on the best course of action

The result is that the wind energy operators can better understand what is happening in the field, plan ahead of time, and properly predict extended operating life (effect of mitigation). All resulting in reduced maintenance cost and improved performance.

**OUTCOMES/CONCLUSION**

This talk will discuss current trends and challenges of Big Data analytics at large from various perspectives. Then, this talk will, describe enabling technologies in Big Data and Analytics in the industrial internet to drive smart decision making and describe the opportunities and challenges of this new paradigm to address wind farm Operations and Maintenance (O&M) expenses and move from reactive to proactive O&M. To conclude, it will outline remaining challenges to be addressed in:

- Integrating large volumes and verity of data to find new insights
- Combining data driven and domain based analytics to get smarter insights
- Providing more enabling infrastructure/platforms for Big data advanced Analytics in the Industrial Internet

**REFERENCES**

[1] https://www.theguardian.com/sustainable-business/2016/dec/06/google-renewable-energy-target-solar-wind-power

[2] https://www.theguardian.com/environement/2016/dec/06/google-powered-100-percent-renewable-energy-2017

[3] https://www.investopedia.com/news/microsoft-makes-largest-wind-energy-investement-msft/?lgl=rira-baseline-vertical

[4] https://techcrunch.com/2016/12/06/google-says-it-will-hit-100-renewable-energy-by-2017/

[5] David Floyer, “Defining and Sizing the Industrial Internet,” Wikibon, June 27, 2013;

[6] Peter C. Evans and Marco Annunziata, “General Electric: Industrial Internet, Pushing the Boundaries of Minds and Machines,” November 2012.

[7] https://www.ge.com/digital/products/asset-performance-management-for-power

[8] Burr Settles, “Active Learning Literature Survey”, Computer Science Technical Report 1648, University of Wisconson-Madison, 2010.

[9] J. Heaton, “An empirical analysis of feature engineering for predictive modeling”, IEEE Southeastcon, 2016