# BIG DATA IN CHEMICAL ENGINEERING

**Matthew N. O. Sadiku[1], Sarhan M. Musa[1] and  Osama M. Musa[2]**

[1]Roy G. Perry College of Engineering

Prairie View A&M University

Prairie View, TX 77446

[2]Ashland Inc.

Bridgewater, NJ 08807

USA

_____

## ABSTRACT

*The big data era is driven by the virtual explosion of new data generated in all fields. The field of big data and analytics is evolving quickly, and an increasing number of tools are being designed for use by chemical engineers. Through big data and analytics, chemical engineers have been able to invent new processes, increase manufacturing efficiency, and cooperate with other industries. This paper is a brief introduction to the use of big data in chemical engineering.*

**Key Words:**  *Big Data, Chemical Engineering, Chemical Industry.*

_____

## 1.  INTRODUCTION

Today companies are generating more data than ever. All branches of engineering, including chemical engineering, are being transformed by big data. Big data refers to data sets with sizes beyond the ability of commonly used software tools to capture, manage, store, and analyze. Sources of big data include multimedia, social media, and Internet of things. Like other industries, chemical companies are using big data to innovate and improve operations. Chemical engineers, whether in industry or in academia, are being asked to manipulate, transform, and analyze complex data sets. Big data tools are already having a great impact in chemical engineering. Big data analytics are now being used by leading chemical companies to recognize patterns, to enhance manufacturing processes, improve marketing, and support innovation. It has been suggested that the traditional three pillars of a modern business — people, process and technology — should now be increased to five pillars by adding data and analytics [1].

Big data analytics is the process of turning data into insights for more informed business decisions. It involves the ability to gain insight from data by applying statistics, simulation, machine learning or other techniques to help organizations make better decisions. Analytical tools let chemical companies go beyond that data and extract the valuable insight.  The insight derived from big data and analytics is crucial to realizing a dynamic ability to understand your business. Due to sound mathematics and computer skills that many chemical engineers possess, the adoption of these tools should prove to be straightforward [2].

## 2.  CHARACTERISTICS OF BIG DATA

Big data describes a collection of data sets of extreme size (e.g. terabytes, petabytes) that are too large and complex to be processed by traditional data analysis tools [3]. There are five properties that can be used to describe big data.

_____

- *Volume:* This refers to large amounts of data generated from devices. It determines the volume of memory required to store the data. The amount of global digital data is doubling every two years. The volume of some commercial data bases may reach many petabytes (1 PB = 1015 Byte).
- *Variety:* This refers to heterogeneity of data types. Data types include numbers, text, image, audio, and video data. Today, data can be structured or unstructured. With the emergence of Industry 4.0 and big data, structured and unstructured data will become increasingly available.
- *Velocity:* This refers to the speed at which new data is generated. New data are arriving faster than we can consume it. There is the need for faster collection, organization and pruning, and processing speed to deal with the large volume of data.
- *Veracity:* This property deals with the truthfulness and accuracy of the data generated. It is difficult to be certain of the veracity of big data. This follows the "garbage in, garbage out" principle.
- *Value:* This is arguably the most important property. Data is often processed to obtain business value. Distributed files are associated with big data as software tools to derive value from big data. As the quality of data improves, so does the value of the insight generated by sophisticated analytics solutions.

## 3. APPLICATIONS

Big data analytics has been applied in fields such as healthcare, aviation industry, finance, energy industry, and supply chain management. Within the chemical industry, the application of big data analytics includes manufacturing, drug discovery, and process industry.

- *Manufacturing*: Big data is one of the key enabling technologies in smart manufacturing. Manufacturing companies own several manufacturing chemical plants operating the same process in different environments and at different locations worldwide. Manufacturing facilities such as chemical plants have been recognized as having components that have complex interactions. The use of big data in chemical manufacturing is helping companies to improve productivity, process efficiency, and product quality across their manufacturing networks [4].
- *Drug discovery:* The role of drug discovery is to develop effective and safe treatments for ailments where there is a significant unmet medical need. Manipulation of proteins in cells and whole organisms is considered standard practice in drug discovery. Technologies using big data in drug discovery are rapidly evolving, Drug discovery is changing, and it is foreseen that artificial intelligence models will soon help research scientists in this endeavor [5].
- *Process industry:* The chemical process industry is one of the earliest adopters of computer-based control. It covers large range spanning commodity chemicals, petrochemicals to consumer products. The large amount of data in industrial processes is gathered at several layers of the automation and for different purposes [6].

## 4. BENEFITS AND CHALLENGES

The incorporation of big data into chemical engineering has already yielded multiple benefits and helped chemical industries gain a competitive advantage. Big data is being leveraged to increase margins, optimize production, enable innovations, and produce highest returns.

Understanding big data is still a challenge. One of the challenges in leveraging big data is the sheer size of the data sets. The generation, curation, and storage of big data have an associated cost. Chemical engineers may need to include big data and analytics in their training so that they are equipped with a variety of analytics tools.

The rise of big data also raises fundamental challenges in privacy, security, and data ownership. Concerns are being expressed over the impact that collecting, storing, and processing large amount of data could have on security. Security of big data is a primary concern in many applications [7]. Analyzing massive amounts of data available in the Internet can impinge on our privacy. Inappropriate analysis of the data can lead to misleading conclusions [8].

## 5. CONCLUSIONS

Both industry and academic partners share high expectations from big data in chemical engineering. Big data is crucial to future success in any industry, particularly chemical industry. It is influencing the chemical industry in a big way. It is time to introduce these developments into undergraduate chemical engineering education. To meet the growing demand for data scientists, over 70 universities in the United States are offering one-to two-year master's degree programs in various related areas such as data science, data analytics, data engineering, and computational science [9].

## REFERENCES

[1] Big data means big opportunities for chemical companies," July 2016 https://home.kpmg.com/ xx/en/home/insights/2016/07/big-data-means-big-opportunities-chemical-companies.html

[2] D. A. C. Beck et al., "Data science: Accelerating innovation and discovery in chemical engineering," AIChE Journal, vol. 62, no. 5, May 2016, pp. 1402-1416.

[3] M. N.O. Sadiku, M. Tembely, and S.M. Musa," Big data: An introduction for engineers," Journal of Scientific and Engineering Research, vol. 3, no. 2, 2016, pp. 106-108.

[4] M. Nino et al., "Requirements for a big data capturing and integration architecture in a distributed manufacturing scenario," Proceedings of IEEE 14th International Conference on Industrial Informatics, July 2016, pp. 1326-1329.

[5] N. Brown et al., "Big data in drug discovery," Progress in Medicinal Chemistry, vol. 5, 2018, pp. 277-356.

[6] L. Chiang, B. Lu, and I. Castillo, "Big data analytics in chemical engineering," Annual Review of Chemical and Biomolecular Engineering, vol. 8, 2017, pp. 63-85.

[7] M. N. O. Sadiku, S. M. Musa, and O. M. Musa, "Big data in the chemical industry," International Journal of Advances in Scientific Research and Engineering, vol. 3, no. 10, Nov. 2017, pp. 20-23.

[8] V. Rajaraman, "Big data analytics," Resonance, August 2016, pp. 695-716.

[9] L. Chiang, B. Lu, and I. Castillo, "Big data analytics in chemical engineering," The Annual Review of Chemical and Biomolecular Engineering, vol. 8, February 2017, pp. 63-85.

## AUTHORS

**Matthew N.O. Sadiku** is a professor in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a fellow of IEEE.

**Sarhan M. Musa** is a professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.

**Osama M. Musa** is currently Senior Vice President and Chief Technology Officer for Ashland Inc. Dr. Musa also serves as a member of the Advisory Board at Manhattan College's Department of Electrical and Computer Engineering as well as a member of the Board of Trustees at Chemists' Club of NYC. Additionally, he sits on the Advisory Board of the International Journal of Humanitarian Technology (IJHT).