Big Data Technologies and Pharmaceutical Manufacturing

B

Joseph E. Kasten
Pennsylvania State University, York, USA

INTRODUCTION

As this article is being prepared, the Covid-19 pandemic continues to be a worldwide problem, although some progress is being made in the world of therapeutics and vaccines. The people of the world have looked to the pharmaceutical researchers, manufacturers, and the logistics organizations that will bring them to the point of use to help us tame this virus and help us all to return to the lives we once lived. The pandemic has brought the pharmaceutical and logistics industries into a focus that they have rarely seen. The focus of this paper is on a very narrow slice of the value chain between the researchers who develop these therapies and those who administer them. This paper examines how the manufacturing of pharmaceuticals benefits from the application of Big Data technologies. More specifically, the paper will describe which pharmaceutical manufacturing processes, as defined in this study, are big data technologies being utilized and how long have the use of these tools to support these processes been studied? Even if the current preoccupation is to get vaccines into the arms of our citizens, an understanding of how these advanced analysis tools promote the efficient and accurate manufacture of drugs is equally important.

The importance of creating a deeper understanding of the role of these technologies in the manufacture of pharmaceuticals is two-fold. First, a better understanding of the analytical tools applied in these operations can provide a platform from which research to identify even better tools and processes might spring. Secondly, an identification of the advanced tools used by pharmaceutical firms in the creation of these amazing compounds will help to broaden the ability of practitioners as well as researchers to identify opportunities to leverage these tools to applications in other related, and possibly unrelated, endeavors. With these benefits in mind, this effort centers on the identification of research that focuses on the technologies in use by the pharmaceutical manufacturers as well as researchers in the field. Using a systematic literature review methodology, the project seeks to identify and categorize the current and future uses of big data technologies in pharmaceutical manufacturing processes.

It is appropriate at this point to define the term "big data technologies." At this point in its development and use, most understand the basic concept of big data: high volume, high velocity, high value, high variability, and often low veracity. These criteria (the five V's) are often used to characterize a big data environment. However, in reality, any dataset with substantial size can fall into a big data category. The issue at hand in this project is not so much the size of the dataset, but rather the functionalities that go along with the concept of big data. These technologies are those that often utilize large datasets to either train the algorithms (such as machine learning and neural networks) or depend on the large dataset to describe a situation that will be used to develop conclusions or evaluate certain patterns (such as in a datamining environment).

The rest of the paper is organized thus: The Background section provides further discussion of the various technologies included in the definition of "big data technologies" as used in this paper. The Methodology section explains how the systematic review process is used in this project. The results

DOI: 10.4018/978-1-7998-9220-5.ch011

of the review are presented in the Findings section. These results are put in their proper context in the Solutions and Recommendations section with some closing thoughts and directions for future research being resident in the Conclusion section.

BACKGROUND

This section provides more specific definitions for the concepts under study in this paper. The first is the term "pharmaceutical manufacturing." For the purposes of the current study, this term applies to the processes used to actually create the drugs and compounds in whatever form they might take (tablets, liquids, etc.) as well as the ancillary operations that are necessary to the success of these processes. These might include, but are not limited to, quality control, safety monitoring, material procurement, etc. In short, whatever aspects of pharmaceutical manufacturing that is returned using the search terms detailed in the next section will be considered. It is not the purpose of this study to decide which activities are considered to be part of the manufacturing process, but rather to report on what the researchers and the practitioners in the field consider them to be.

The next term to consider is what has already been designated "big data." As mentioned in the Introduction, most can identify big data when it is presented and likely the technologies associated with it. However, a more specific listing of these technologies is necessary to properly review the literature:

- Deep Learning: A subtype of machine learning that refers to tools such as Artificial Neural Networks (ANN). These tools are trained to make decisions such as classifications or speech recognition. These are considered to be related to big data because the datasets required to train and verify the algorithms commonly fall into the general definition of big data.
- Machine Learning: A broad category of algorithms that can be trained to create the ability to make decisions. There are many different types of training methods and architectures, a full discussion of which is out of the scope of this paper.
- Data Analytics: A broad term covering all of the statistical and mathematical tools and techniques commonly used to analyze and draw meaningful information from big data repositories.
- Datamining: A category of tools that are able to sift through large and sometimes disparate datasets to find trends or categorize the data into subsets that are more descriptive of the concepts buried therein.

This is by no means an exhaustive list of technologies and tools associated with big data, but it provides a representative list of categories of those tools that might be found in a manufacturing environment of this type. In many cases, the tools used to search the databases used in this study have broad enough thesauri that the tools not specified in the search are still returned as part of the search results.

METHODOLOGY

The methodology known as a systematic review of the literature requires that a well-defined, easy to replicate method be in place. Systematic reviews have been used successfully to provide a comprehensive overview of the literature describing a specific area of inquiry. In the case of pharmaceutical manufacturing, this methodology has been used to draw a number of important conclusions about the