

How Data Science Drives Innovation in Pharmaceutical Clinical Trials

Big data has transformed how we manage, analyze, and apply data across industries. Healthcare, especially, is a notable area where data analytics and data science make a significant impact. Currently, data science is spearheading many innovations in healthcare, particularly in pharmaceutical clinical trials.

Along with the emergence of Artificial Intelligence (AI) and Machine Learning (ML) technologies in the past five years, data science unlocks new possibilities in pharmaceutical research and development. AI/ML tools can churn through massive datasets and provide accurate results in just minutes, uncovering valuable insights that would take an enormous number of hours for humans to perform. Here are some key benefits that data science can bring to the drug development process:

More Accurate Clinical Trials

A large pharmaceutical company can have thousands of ongoing clinical trials with millions of datasets. With so many data points, the need for effective data management and data analysis is greater than before. Mismanaged data can lead to costly mistakes that waste precious resources and staff time, or worse, put the entire clinical trial at risk.

Despite the risks, many clinical studies still rely on basic data collection and verification methods, such as counting leftover pills in bottles manually, sending patient medical records via fax, and using patients' paper diaries to determine their medication adherence. Often, these tasks fall on the patient, who is more likely to forget or make mistakes.

By using advanced technologies like AI, millions of records can be searched for pertinent information. This data enrichment process is impossible for humans to do, but it can be done in real-time and update consistently with data science practices. The result is more accurate and validated data, which lays the foundation for more precise trials.

Safer Production of Drugs

Manufacturing a new pharmaceutical drug is a risky process with many variables. However, by applying machine learning and statistical techniques, the outcomes of randomized clinical trials for new drug candidates can be forecast through predictive modeling. The result is more accurate and timely estimates of the risks and rewards for all stakeholders involved, including researchers, regulators, and the patients themselves.

By applying data science techniques to Big Data collected during the drug discovery process, scientists can simulate the reaction of drug-using body proteins and various types of cells and conditions. The

resulting drug from that clinical trial is far more likely to be approved by the Food and Drug Administration, as well as cure diverse patient profiles.

With more accurate measures of the risk of drug development, researchers can design improved clinical trials that eliminate costly delays in a market launch. Data science and analytics also expand the criteria for patient selection. Factors like individual characteristics, disease status, and genetic include allowing trials to only target patients who match the inclusion and exclusion criteria.

More Efficient Trials

Not only does data analytics lead to more informed decision-making during the drug development process, but it can also improve the efficiency of research and clinical trials. Predictive modeling of drugs and biological processes will become more refined, which can help identify new potential candidate molecules that can be successfully developed into drugs with a high degree of certainty.

By leveraging big data and automation solutions, pharmaceutical companies can respond in real-time to emerging insights from the clinical data. They can also be more efficient in their trials by running smaller tests of equivalent power or shorter trial times. These small efficiencies quickly compound and can reduce the trial time by a factor of months or even years.

As data continue to pour in from thousands of clinical trials, evolved data strategies will make the drug-development pipeline more efficient. For example, the widespread use of electronic data such as electronic medical records can reduce the likelihood of data errors due to manual or duplicate entry.

New Data Analytics, New Tools

Despite the rise of big data, many organizations are still relying on old technologies for data collection and analysis. For years, Microsoft Excel has been the preferred data analysis tool for professionals in nearly every setting, including clinical research and trials during drug development. While Excel comes with numerous useful tools and spreadsheet functions, its lack of “project management” capabilities makes it less ideal for handling the data quantity in clinical trials today.

To take advantage of big data analytics, organizations have turned to alternatives such as R and Python. R is a free and open-source programming language that offers advanced data analysis functions and capabilities. Compared to Excel, R programming can handle larger data sets and generate more detailed visualizations. With clinical trial data, R can streamline complex processes to reduce error and reliably reproduce results.

Whereas R is solely used for data analysis, Python is another programmable language that can develop system backends, web pages, or even games. In Python, data scientists can manipulate data like functions in Excel. However, they can also make use of advanced statistics and machine learning capabilities. Python can scale and work with larger and multiple datasets better than Excel, which is unsuited as a data storage for thousands of patient entries and millions of tables. Finally, Python can automate much of the data reports and analyses to save hours of manual work.

Conclusion

Taking a new pharmaceutical drug to market is a slow and costly process with frequent roadblocks, but big data and data science have reduced risks in drug development. Thanks to improved data strategies and AI/ML solutions, organizations can expedite life-saving drugs through clinical trials in the drug discovery process. However, you need trained data scientists that understand algorithms, data collection, storage, validation, and visualization to generate valuable insights.

Data Science is a new and constantly evolving field, but with effective training, your organization can develop a skilled data science team and upskill existing employees. For example, your employees can quickly learn new languages like R and Python to minimize the time spent culling and wrangling data. Not only does data science training prepare your organization to handle Big Data, but it also creates significant competitive advantages during drug development.