A Primer on Data Science in Healthcare

Chandan Marella, Lead Data Scientist

You've probably heard of data science by now. Perhaps you've seen the term in a health IT headline or perhaps your health system has data scientists on staff. Healthcare, with all its longitudinal data over a patient's lifetime, holds great potential to take advantage of what data science has to offer.

Naturally, data science plays a critical role in our work with healthcare organizations across the country. Seeing as it is still a new field, we want to give a primer on data science and the opportunities it presents to improving healthcare delivery in the U.S.

Why Healthcare Needs Data Science

First, what is data science? It is a field that uses scientific methods, processes, algorithms, and systems to extract insights from data. Data science applies to data that is structured, meaning it is organized in an intentional and consistent way for easy searching in a database. Data science also applies to unstructured data, which has no organizational parameters and is generally more difficult to collect, process, and analyze.

Next, how might data science be used to transform healthcare delivery? A Ponemon Institute survey revealed that healthcare stores about 30 percent of global data, with the U.S. accounting for approximately 1.2 billion clinical records each year. With this magnitude of data waiting to be mined and explored by data scientists, the possibilities seem endless. Here are a few practical applications that we, at Ancore Health, believe should be further expanded and embraced by healthcare leaders.

- **Improving the patient scheduling process:** Predictive analytics can be used to forestall appointments that are no shows. Triage and scheduling could be supported by efficient data science algorithms, which could identify patients that are likely to skip an appointment without prior notice.
- **Predicting population health trends and risk scores for chronic diseases:** Data science can help predict the costs associated with a particular population based on their disease burden, while also identifying individuals with an elevated risk of developing chronic conditions earlier in the disease's progression. Data science can also help with target prevention therapies by creating risk scores based on lab testing, biometric data, claims data, patient-generated health data. By forecasting population health and targeting patient screening, mature health systems could save up to 10 percent of total care costs, and providers transitioning to value-based care can improve risk management.

- **Avoiding 30-day hospital readmissions:** Currently, one-in-five Medicare patients is readmitted to a hospital within 30 days of discharge. Three quarters of these readmissions (75%) are actually considered avoidable. Data science can help predict these preventable readmissions, which would help the health system avoid readmission penalties, improve performance against bundled payments, lower the cost of care, and improve the patient experience.
- **Delivering preventive care and personalized treatments:** Only 9 percent of the total US health expenditure is related to prevention. Data science and big data techniques can help create interventions that prevent diseases from occurring and support early treatments. Data science can also help come up with personalized treatment for individuals based on their disease subtype, risk, prognosis, or treatment response.

Healthcare Lagging in Adaptation on Data Science

In comparison to other sectors, healthcare is still lagging behind in its adoption of data science. The pharmaceutical industry has already begun using data science to optimize the clinical trials pipeline and understand potential market demands. However, according to research by Mckinsey Global Institute, the US has gained no more than 10-20 percent of the opportunity offered by data science and artificial intelligence (AI) in healthcare. Retail and location-based services are capturing between 30 to 60 percent of data science opportunities.

Healthcare policies can sometimes become an obstacle to the industry's ability to fully leverage data science. We at Ancore Health see this playing out in three areas. The first is with the collection, organization, and storage of these large datasets. Healthcare organizations must create systems that can handle these large datasets, integrate them, and ensure interoperability with various sources.

The second is the security needed to protect highly sensitive healthcare data, required by the Health Insurance Portability and Accountability Act (HIPAA) of 1996. Healthcare providers must create systems that adhere to strict regulations and protect patient data while also enabling interoperability with other sources, which can be exceedingly difficult.

The third area is attracting the talent needed to design these complex systems. This delay is likely due, in part, to the relatively fragmented nature of the healthcare sector and the economic realities of public healthcare systems.

The use of data science in other sectors shows that these challenges can be overcome with sufficient leadership and commitment from policymakers.

If you have questions about finding the right data strategy for your organization's goals, contact us at:

(615) 873-0571

contactus@ancorehealth.com