BIG DATA

IN PRACTICE

HOW 45 SUCCESSFUL COMPANIES USED BIG DATA ANALYTICS TO DELIVER EXTRAORDINARY RESULTS

BERNARD MARR

WILEY

6
APIXIO

How Big Data Is Transforming Healthcare

Background

California-based cognitive computing firm Apixio were founded in 2009 with the vision of uncovering and making accessible clinical knowledge from digitized medical records, in order to improve healthcare decision making. With their team of healthcare experts, data scientists, engineers and product experts, the company have now set their sights on enabling healthcare providers to learn from *practice*-based evidence to individually tailor care.

What Problem Is Big Data Helping To Solve?

A staggering 80% of medical and clinical information about patients is formed of unstructured data, such as written physician notes. As Apixio CEO Darren Schulte explains, "If we want to learn how to better care for individuals and understand more about the health of the population as a whole, we need to be able to mine unstructured data for insights." Thus, the problem in healthcare is not lack of data, but the unstructured nature of its data: the many, many different formats and templates that healthcare providers use, and the numerous different systems that house this information. To tackle this problem,

Apixio devised a way to access and make sense of that clinical information.

How Is Big Data Used In Practice?

Electronic health records (EHRs) have been around for a while, but they are not designed to facilitate data analysis and contain data stored across a number of different systems and formats. So before Apixio can even analyse any data, they first have to extract the data from these various sources (which may include doctors' notes, hospital records, government Medicare records, etc.). Then they need to turn that information into something that computers can analyse. Clinician notes can come in many different formats – some are handwritten and some are in a scanned PDF file format – so Apixio use OCR (optical character recognition) technology to create a textual representation of that information that computers can read and understand.

Apixio work with the data using a variety of different methodologies and algorithms that are machine learning based and have natural language-processing capabilities. The data can be analysed at an individual level to create a patient data model, and it can also be aggregated across the population in order to derive larger insights around the disease prevalence, treatment patterns, etc. Schulte explains, "We create a 'patient object', essentially a profile assembled using data derived by text processing and mining text and coded healthcare data. By creating this individual profile and grouping together individuals with similar profiles, we can answer questions about what works and what doesn't in those individuals, which becomes the basis for personalized medicine."

Traditional evidence-based medicine is largely based upon studies with methodological flaws, or randomized clinical trials with relatively small populations that may not generalize well outside that

particular study. By mining the world of practice-based clinical data – who has what condition, what treatments are working, etc. – organizations can learn a lot about the way they care for individuals. Schulte, a physician who was Apixio's chief medical officer before being appointed CEO, says: "I think this could positively disrupt what we [in the healthcare industry] do. We can learn more from the *practice* of medicine and refine our approach to clinical care. This gets us closer to a 'learning healthcare system'. Our thinking on what actually works and what doesn't is updated with evidence from the real-world data."

The first product to come from Apixio's technology platform is called the HCC Profiler. The customers for this product fall into two groups: insurance plans and healthcare delivery networks (including hospitals and clinics). Medicare forms a big part of their business, especially those individuals within Medicare who have opted into health maintenance organization (HMO) style plans (called Medicare Advantage Plans), which accounted for nearly 17 million individuals in the US in 2015. Health plans and physician organizations have an incentive to manage the total cost of care for these individuals. To do this, these organizations need to know much more about each individual: What are the diseases being actively treated? What is the severity of their illness? What are various treatments provided to these individuals? This is much easier to understand when you can access and mine that 80% of medical data previously unavailable for analysis, in addition to coded data found in the electronic record and in billing or administrative datasets.

What Were The Results?

For those patients in Medicare Advantage Plans, Medicare pays a "capitated payment" to the sponsoring health plan or provider organization – a monthly payment calculated for each individual based upon the expected healthcare costs over that year. The payment is

calculated using a cost prediction model that takes into account many factors, including the number, type and severity of conditions treated for the individual. Understanding these conditions is critical not just for estimating the cost of healthcare for individuals over a given period but also because the information is also very useful to help better manage care across the population. Traditionally, in order to understand such patient information, experts trained in reading charts and coding the information ("coders") would have to read the entire patient chart searching for documentation related to diseases and treatment. This is a laborious and expensive way of extracting information from patient records, and one that is fraught with human error. Apixio have demonstrated that computers can enable coders to read two or three times as many charts per hour than manual review alone. In addition to speeding up the process of chart review, Apixio have found that computers are more accurate as well. The accuracy improvement can be as high as 20% relative to what a coder manually reading the chart would be able to find themselves.

An additional benefit is the computer's ability to find gaps in patient documentation, defined as a physician notation of a chronic disease in the patient history without a recent assessment or plan. For example, over a nine-month period within a population of 25,000 patients, Apixio found over 5000 instances of diseases that were not documented clearly and appropriately. Gaps like this can lead to an inaccurate picture of disease prevalence and treatment, which can negatively affect the coordination and management of patient care. These document gaps provide a great way to better educate physicians on proper documentation. Schulte explains: "If you don't get that information right, how can the system coordinate and manage care for the individual? If you don't know what it is you're treating and who's afflicted with what, you don't know how to coordinate [care] across the population and manage it to reduce costs and improve the outcomes for individuals."

What Data Was Used?

Apixio work with both structured and unstructured data, although the bulk of their data is unstructured, typewritten clinical charts. This can include GP notes, consultant notes, radiology notes, pathology results, discharge notes from a hospital, etc. They also work with information on diseases and procedures that are reported to the government (in this case Medicare).

What Are The Technical Details?

Apixio's Big Data infrastructure is composed of well-known infrastructure components, which include non-relational database technology like Cassandra and distributed computing platforms like Hadoop and Spark. Apixio have added to this their own bespoke orchestration and management layer that automates a system that cannot be operated manually at the scale Apixio operate at. Everything is operated on Amazon Web Services (AWS) in the cloud, which Apixio selected for its robustness as well as its healthcare privacy and security and regulatory compliance. Everything is processed and analysed in-house using their own algorithms and machine-learning processes, as opposed to working with an external Big Data provider. Apixio created their own "knowledge graph" to recognize millions of healthcare concepts and terms and understand the relationships between them. That type of tool is healthcare-specific: an out-of-the-box solution from a Big Data provider working across a range of industries just wouldn't work for them. Patient charts in PDF or TIFF files are the primary data provided by health insurance plans, given their process for acquiring charts from provider offices via faxing, or printing and scanning the requested records in the medical office. Therefore, Apixio developed sophisticated technology to leverage and scale OCR to make machine-scanned medical charts readable by their algorithms. Sophisticated computational workflows that pre-process images, set parameters in the OCR engine and correct

output had to be developed to extract the text available in a scanned chart.

Any Challenges That Had To Be Overcome?

Getting healthcare providers and health insurance plans to share data is a real challenge, which holds back attempts to assemble large datasets for Big Data-driven knowledge generation. Apixio overcame these hurdles by demonstrating that they were offering real value. "Our value proposition is strong enough to overcome any trepidation about sharing this data … unless you solve a real critical problem today, none of these organizations will give you access to any real amount of data," Schulte explains.

Which brings us to the next challenge: data security. Thanks to some high-profile health data breaches, security is a hot topic in this field. For Apixio, the importance of data security and their legal requirements were a top consideration from the get-go. Schulte refers to data security as "table stakes", meaning it is an essential requirement for anyone wanting to operate in the healthcare Big Data arena. "For every new contract we have to demonstrate our security. And being on AWS certainly does help in that regard … it takes a good amount of their anxiety away," he explains. Patient data must be encrypted at rest and in transit, and Apixio never expose personal health information (PHI) unless access is absolutely needed by Apixio staff. "The proof is in the pudding," Schulte says. "Large health insurance plans would not sign contracts and do business with us if they didn't feel it was secure enough to do so."

What Are The Key Learning Points And Takeaways?

Big Data in healthcare is still in its infancy, and there remains a lot of hype around the possibilities, sometimes at the expense of tangible results. Schulte confirms this: "CIOs at hospitals don't often see a lot

of problems actually being solved using Big Data. They see a lot of slick dashboards which are not very helpful to them. What's helpful is actively solving problems today … for example, ensuring appropriate care and reducing costly, ineffective treatments … It's important to focus on actual results, something tangible that has been accomplished. Don't just lead with, 'Hey, I created some little whizz-bang data science tool to play with.'" This emphasis on results and outcomes is just as essential for business as it is for the healthcare industry.

If you strip away the hype, though, it's still clear that we're on the verge of exciting changes in the way we understand, treat and prevent diseases. Schulte agrees: "We're in a new world in terms of the way healthcare is going to be practised, based upon these data-driven insights, and Big Data is a way of helping us get there."

REFERENCES AND FURTHER READING

Find out more about Apixio and Big Data at:

http://www.apixio.com

http://www.smartdatacollective.com/bernardmarr/339832/why-we-need-more-big-data-doctors

WHAT'S NEXT:

If you enjoyed the insights from these fascinating case studies, you're going to love the 45 additional case studies I present in *Big Data in Practice*.

If you want more information about the SMART formula for Big Data, I explain it in much more detail in my previous book, *Big Data: Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance*.
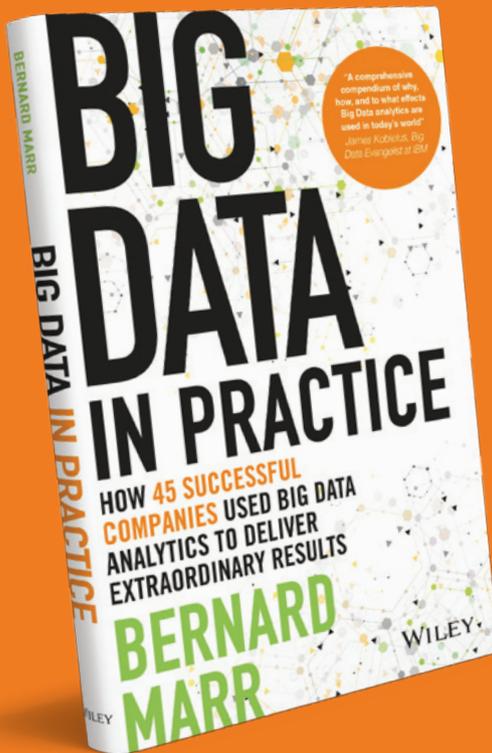
ABOUT THE AUTHOR

Bernard Marr is a leading global authority and best-selling author on organizational performance and business success. He is a LinkedIn Influencer, he writes the Big Data Guru blog and is the world's #1 expert on big data. He regularly advises leading companies, organizations and governments across the globe, and is an acclaimed and award-winning keynote speaker, researcher, consultant and teacher.

Bernard is the founder and CEO of the Advanced Performance Institute. Prior to this he held influential positions at the University of Cambridge and at Cranfield School of Management. Bernard Marr's expert comments on organizational performance have been published in *The Times*, *The Financial Times*, *The Sunday Times*, *Financial Management*, the *CFO Magazine* and *The Wall Street Journal*.

Connect with Bernard