Penn Signals Big Data Analytics Helps Penn Medicine Improve Patient Care

Industry

Healthcare

> "With open source and cloud-based solutions like the Trusted Analytics Platform, we have the potential to provide our clinicians and data scientists with a platform for development and testing where we can quickly build, explore, and deploy our analytic applications."
>
> – Mike Draugelis
> Chief Data Scienctist
> Penn Medicine

Overview

Compared to other industries, healthcare as a whole has been a late adopter of big data predictive analytics. This may be due in part because of concerns about patient confidentiality and fears of security breaches on open source Hadoop-based systems. Demonstrating a return on investment (ROI) for big data solutions to hospitals and healthcare providers can also be a challenge.

Ironically, most healthcare organizations have an abundance of patient data at their disposal that they could use to benefit patient care and medical research while driving down costs…if they could only implement a big data solution.

This paper shows how one healthcare institution—Penn Medicine—is using big data solutions to derive new insights and improve patient care.

Penn Medicine's Big Data Vision – Penn Signals

Penn Medicine is a $4.3 billion organization with more than 2,000 physicians providing services to the Hospital of the University of Pennsylvania, Penn Presbyterian Medical Center, Pennsylvania Hospital, Chester County Hospital, Lancaster General Health, and a health network that serves the city of Philadelphia, the surrounding five-county area, and parts of southern New Jersey.

Considered a leader in healthcare predictive analytics, the Penn Medicine data science team is dedicated to improving patient outcomes through analytics. Specifically, they want to harness the full power of clinical data to help clinicians identify patients at risk of critical illnesses that may have been missed by current diagnostic techniques. In the process, they are also developing solutions that will remove the barriers to developing analytic models and accelerate the deployment of analytic applications based on those models, all of which they intend to share with other health organizations via Open Source.

To accomplish this, Penn Medicine needed a platform for rapid development and deployment of predictive analytics applications that could be applied to detect patients at risk of critical illnesses. The platform they developed is called Penn Signals. Penn Signals is a collaborative data science platform developed by the Penn Medicine data science team that combines clinical data at scale with big

Key Challenges

- Harness clinical data to improve patient risk stratification
- Deliver useful predictive alerts to clinicians
- Accelerate development and delivery of predictive applications

Solution

- Penn Signals is a collaborative data science platform that combines clinical data at scale with advanced analytics to develop and deploy predictive applications.
- Trusted Analytics Platform is an extensible open source platform designed to allow data scientists and application developers to deploy solutions without having to worry about infrastructure procurement or platform setup.

Benefits

- Identified patients at risk of critical illnesses that may have been missed by current diagnostic techniques
- Reduced patient re-admissions through early risk stratification and intervention
- Improved patient care and well-being
- Accelerated development of predictive applications

data to allow researchers to explore solutions, allow developers to develop predictive applications, and provide a platform for deployment. The first applications of Penn Signals focused on sepsis and heart failure.

Sepsis and Heart Failure

To gauge the success of Penn Signals, Penn Medicine decided to run clinical pilot case studies involving two very different acute events—sepsis and heart failure. They hoped these pilots would provide tangible improvement to patient care that might convince other healthcare institutions of the viability of such a big data solution.

According to the Centers for Disease Control (CDC), sepsis (blood infection) affects more than a million Americans annually and is the ninth leading cause of disease-related deaths and the #1 cause of deaths in intensive care units.[1] Worldwide incidents exceed 20 million cases a year, and mortality due to septic shock may approach 50 percent, even in industrialized countries.[2] The mortality rate is approximately 40 percent in adults and 25 percent in children.[3]

Treatment guidelines call for the administration of broad-spectrum antibiotics within the first hour following recognition of septic shock. Prompt antimicrobial therapy is critically important, as the risk of dying increases by approximately 10 percent for every hour of delay in receiving antibiotics.[4]

Traditional methods of sepsis identification generally only detect about half of the cases, and even then detection typically occurs just two hours before the patient succumbs to septic shock. A diagnostic process that could help staff identify sepsis earlier could profoundly improve treatment success and lower the mortality rate.

Heart failure (the inability of the heart to pump enough blood to meet the needs of the body and lungs) is amazingly common, affecting 5.8 million people in the United States[5] and about two percent of adults worldwide.[6]

It is the number one reason that people in the developed world—especially those over 65—are admitted to the hospital.[7]

Even though treatment itself is relatively inexpensive and typically involves lifestyle modifications (reduced smoking, physical exercise, dietary changes), as well as medications, treatment costs are high, mostly because of hospitalization, with estimates exceeding $35 billion per year in the United States.[8] Exacerbating those costs is the fact that a quarter of patients hospitalized in the United States are readmitted within 30 days, and half seek readmission within six months after treatment.[9] Penn Medicine deduced that an improved algorithm that could correctly identify patients and get them on the right regimen earlier could reduce treatment costs, reduce hospital readmissions, and improve patient care.

The Solution

Like many hospital groups, Penn Medicine has a wealth of patient data. Their clinical data warehouse currently holds records on three million patients stretching back more than 10 years. In 2013, the Penn Medicine data science staff envisioned a big data environment that would use all of this data to improve their clinical pathways and predictions. They pooled thousands of variables together—vitals, lab results, medications—and fed these variables into a real-time matrix that they could apply algorithms in order to make predictions.

For the sepsis pilot, which involved 150 of Penn Medicine's 1200 to 1500 total patients, Penn Medicine was able to correctly identify about 85 percent of sepsis cases (up from 50 percent), and made such identifications as much as 30 hours before onset of septic shock (as opposed to just two hours prior, using traditional identification methods).

The first problem for the heart failure pilot was patient identification, so the analytics team started by creating a heart failure detection algorithm.

Heart failure patients who are not diagnosed on their first hospital visit—which may be for a totally unrelated injury or malady—return to the hospital when a negative event related to heart failure does occur later, and often for a prolonged stay. Identifying at-risk patients before they have a negative event can result in better care for the patient and reduced costs for the hospital.

Shortly after deploying this algorithm, the clinical team realized they finally had an accurate measure of the quantity and distribution of heart failure patients within or between the hospitals, perhaps for the first time ever. By their own estimates, between 20 and 30 percent of heart failure patients had not been getting properly identified using standard diagnostic tools. With the Penn Signals predictive model, they were identifying those patients and giving them the education to self-manage their condition successfully. As a result, readmission rates dropped.

Embedding Data Scientists With Clinical Teams

Penn Medicine sees great bidirectional benefits and has already witnessed measurable, positive change from seating data scientists and clinicians at the same table. It is important to pair the data science and healthcare expert so the two groups can iteratively help nudge each other to new insights.

Clinical experts, who know what the important variables are, can point data scientists in the right direction. Then the data scientists can use machine learning techniques to find relations that the clinicians never realized, at scale and in real-time.

The pilot programs confirmed that the variables clinicians look at are important in forecasting events such as severe sepsis; however, the data science team was able to find dozens of additional variables that also proved important. These analytical insights provide two benefits:

– First, the hospital gets a more powerful and more accurate forecasting algorithm, which produces better results.

– Second, the new discoveries create new paths of research for clinicians and medical researchers to investigate.

From Insight to Action

Once the data science team was able to produce reliable predictions, their next challenge was to figure out how to deploy this in real time so that a care provider at the point of care can get information when it's most meaningful in a way that they can act on it.
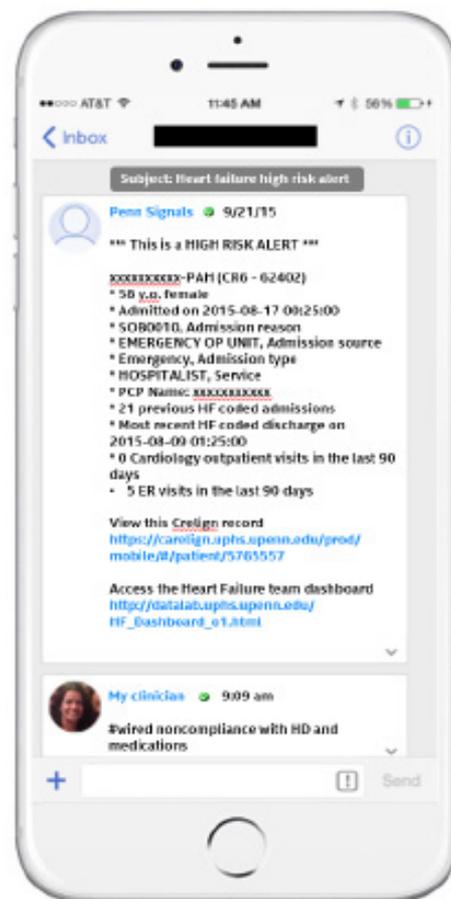

Figure 1

The solution chosen had three components:

- Immediate alerts. A secure text mobile app called Cureatr sends secure text message alerts (see Figure 1) to the smartphones of specific care providers (based on the location of the patient, who is working that day, etc.). These alerts contain the patient's location as well as clinical information that helps the care provider decide whether or not the alert is relevant and what services the patient might need.

- Automated task lists. A set of tasks that the data scientists devise with the care team. These manage workflow in the EMRs and identify patients who are in grave danger.

- Performance dashboard. The combined clinical and data science team created a dashboard that reports how the care team is acting in the care pathway, how to reduce readmission rates, how to improve quality of life for patients, etc. From the data science side, they are able to track the accuracy of an algorithm and could alter the algorithm if they see its accuracy diminish.

Accelerating Analytics Deployments With Trusted Analytics Platform (Tap)

Now that the Penn Medicine data science team has some successful pilots under their belt, the next challenge is how to take those solutions and deploy them in production, in an IT system that has to keep 300 to 3,000 different applications running. They want to take the best solutions that depend on open source technology, that depend on big data techniques, and deploy them in an infrastructure that has to have 24/7 reliability and in a way that

doesn't take months and months to do. Those constraints are pain points that the Penn Medicine data science team is working to solve with the next generation of Penn Signals.

To do this, Penn Medicine is excited about the potential of using TAP to allow clinicians and data scientists to have a development, test, and QA environment, where they can quickly build, quickly explore, and quickly deploy new predictive analytics applications.

TAP (see Figure 2) is an extensible open source platform designed to allow data scientists and application developers to deploy solutions without having to worry about infrastructure procurement or platform setup. TAP empowers small teams by unburdening them from dealing with the infrastructure and the platform, letting them focus on developing applications and connecting with experts and stakeholders outside of the data science sphere.

TAP is effective for building solutions and achieving complex interactions at scale. More importantly, TAP allows data scientists to try a variety of ideas quickly to see which ones work best. Data science is highly iterative, and TAP quickens iteration by providing a variety of possibilities to find the best approach.

Penn Medicine is exploring the use of TAP in joint projects with Intel®.

The goal of this project was to improve Penn Medicine's ability to predict patients who are at risk of being re-admitted to the hospital for heart failure in the next 30 or 90 days of discharge by using a patient's medication history as an added predictor. The Intel® data science team worked with Penn Medicine to use TAP analytics tools to utilize data they had on medications each patient had taken prior to being admitted to the hospital and during their stay.

By treating this information as a "document" and applying machine learning approaches typically used in
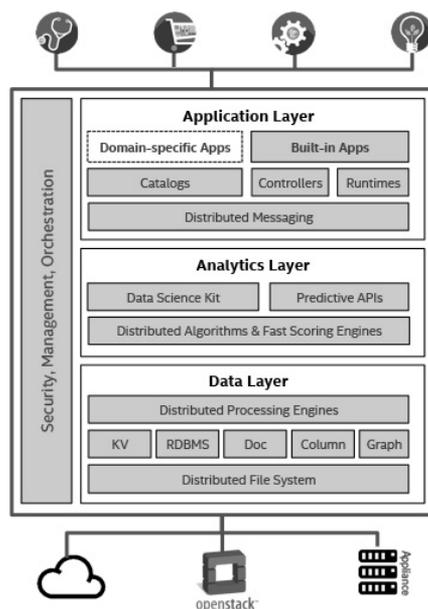

Figure 2

document processing they were able to find topics and categories of illnesses. In medical terminology, this is called phenotyping a patient or categorizing the type of illnesses a patient has, defined by medications alone.

The chart in Figure 3 compares the predictive accuracy of multiple predictive approaches and shows that adding medication history improves predictive accuracy by up to 15 percent for patients being discharged from hospital. These predictions improve the hospital's ability to target additional out-patient care to hopefully keep these patients from being re-admitted.
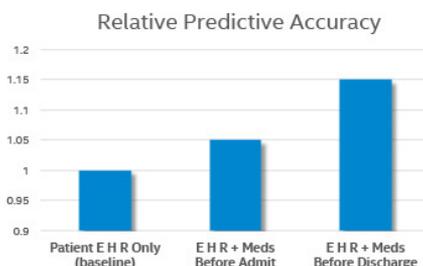

Figure 3

Summary

According to a recent Gartner Group study of nearly 300 respondents, 65 percent cited their inability to identify the value of a big data system as the single greatest barrier that prevents them from adopting their own big data solutions.[10]
This implies a need for education and some guidance to help others make that first step, or at least prove the utility of such a deployment.

The healthcare industry can benefit greatly from the big data insights predictive analysis provides, as Penn Medicine's programs illustrate. Trusted Analytics Platform has the potential to make developing such cloud-based analytics easier for even small data science teams to implement successfully.

To this end, Penn Medicine is excited to share what they have accomplished in Penn Signals with other healthcare systems.What Penn Medicine has seen, particularly in data science, is that cutting edge technologies are often open source projects. Penn Medicine has built on a foundation based on open source and now wants to contribute back to that community. For that reason, their long-term plan is to open source Penn Signals.

Learn More

Trusted Analytics Platform:
www.trustedanalytics.org

Intel Big Data:
www.intel.com/bigdata

Penn Medicine:
http://www.pennmedicine.org/

Endnotes

1. Centers for Disease Control and Prevention. Sepsis questions and answers. May 2014. **http://www.cdc.gov/sepsis/basic/qa.html**

2. PhysOrg. Researchers make blood poisoning breakthrough. June 4, 2010. **http://www.physorg.com/news194879092.html**

3. Huether; McCance; eds. Understanding Pathophysiology (4th ed.). ISBN 9780323049900. 2008.

4. Martin, Greg. Sepsis, severe sepsis, and septic shock: Changes in incidence, pathogens, and outcomes. April 2013.
    http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3488423

5. Bui; Horwich; Fonarow. Epidemiology and risk profile of heart failure. January 2011. **www.ncbi.nlm.nih.gov/pmc/articles/PMC3033496**

6. McMurray; Pfeffer. Heart failure. 2005. **www.ncbi.nlm.nih.gov/pubmed/15924986**

7. Krumholz; Chen; Wang; Vaccarino; Radford; Horwitz. Predictors of readmission among elderly survivors of admission with heart failure. 2000.
    www.ncbi.nlm.nih.gov/pubmed/10618565

8. Rosamond; Flegal; Furie; et al. Heart disease and stroke statistics--2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. January 2008. w**ww.ncbi.nlm.nih.gov/pubmed/18086926**

9. Elixhauser; Steiner. Readmissions to US Hospitals by Diagnosis, 2010. HCUP Statistical Brief #153. Agency for Healthcare Research and Quality. April 2013.
    http://hcup-us.ahrq.gov/reports/statbriefs/sb153.jsp

10. The Gartner Group. Gartner survey analysis. September 2014. Industry participant interviews (N=282).

(intel®)