Predictive Analytics for Healthcare: What's Now, What's New and What's Next

Dr. Michael Zeller of Zementis interviews Dr. Ankur Teredesai of the University of Washington

Ankur Teredesai, Ph.D.
Professor of Computer Science and Systems
University of Washington Tacoma, Institute of Technology

Dr. Teredesai describes some of the groundbreaking work that he and his team at the University of Washington are doing in predictive analytics for healthcare -- not just within the data science lab, but also with patients, providers and policy makers.

MZ: Hello, and welcome to the Zementis video chat on the topic of real-world use cases for predictive analytics. Today it is my pleasure to welcome Dr. Ankur Teredesai. He is here to share his perspectives about predictive analytics, data mining and machine learning, and today with the particular focus on exciting developments in the healthcare industry.

But, before we begin, please allow me to properly introduce Dr. Teredesai: He comes to us from the University of Washington Tacoma, where he is a Professor of Computer Science & Systems at the Institute of Technology. He also heads the Center for Data Science at the University of Washington, and he serves as the Information Officer for ACM[1] SIGKDD (that is the ACM Special Interest Group on Knowledge Discovery and Data Mining[2]) and he is currently an associate editor for ACM SIGKDD Explorations as well as IEEE Transactions on Big Data, and he serves on program committees of major international conferences in data mining, machine learning and related areas.

Dr. Teredesai's research focuses on ways that data science can be applied to crucial societal issues to achieve a positive impact that drives social good – so a very noble cause. Some of his recent applied research actually contributes to risk prediction for readmission due to chronic conditions such as heart failure, while other applications of his work have enabled trust-enhanced recommendations, for example.

Ankur, Welcome!

AT: It's a pleasure to be here, Michael.

MZ: Ankur, given your current focus on the healthcare industry, I thought it would be very interesting if you could share with us your perspectives on a few statistics

[1] *Association for Computing Machinery: http://www.acm.org/*
[2] *http://www.sigkdd.org/*

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878

that I found very interesting if not rather alarming, and I'd ask you to explain to us how you see the role of data science in identifying these situations and understanding root causes, potentially improving patient quality of life and quality of care.

AT: Sure Michael. Last year, for example, nearly 1 in 5 -- one in five! – Medicare patients who had been hospitalized were readmitted within a month. Roughly 2 million patients return within a year, many of them return repeatedly, costing Medicare around $26 billion. Official estimates actually say that $17 billion of that $26 billion is spent on something that we term "potentially avoidable readmissions". That is huge, because nearly one twentieth of the spend that we are experiencing as taxpayers is being spent on something that the medical community considers avoidable.

MZ: It's really shocking to look at those numbers. So, what is your perspective on these alarming numbers, and how can predictive analytics really help reduce the incidence of such events?

AT: It's an opportunity for data science. Data science today has become a catalyst for transforming the practice of healthcare. Consider, for example, Jim. Jim is a 70-year-old heart failure patient with hypertension and complications due to diabetes. Jim unfortunately was hospitalized three times last year, twice within just the last two months, simply because there was no one to remind him to take his medications every day. Usually his daughter would do so, but due to her deadlines at work and other reasons, she was not able to remind him and he got hospitalized several times. Or, take another example of John, who is 55 and has an inactive lifestyle, but simple respiratory exercises would have kept him going for many years. Instead he underwent an expensive surgical procedure that made him even less active. Simply avoidable procedural intervention would have helped him lead an active lifestyle.

Over the course of the last 3 years or so, my group at University of Washington has partnered with leading cardiologists in our region to develop machine learning models, not just for accurately predicting which patients are high-, medium- or low-risk for 30-day readmission, but to actually identify the likelihood of different care pathways that these patients should be recommended, so that their risk scores are lowered and we can address multiple chronic conditions for these patients.

MZ: Very fascinating. So, how does your work at the Center for Data Science advance not just the academic discipline of data science but really achieve a greater social good? And what are some of the examples that you think really illustrate your mission in a powerful way?

AT: So Mike, the UW Center for Data Science was founded by our late chancellor Debra Friedman with the vision to enable data-driven decisions, particularly for domains that are very rich in data but very poor in tools; for example, healthcare or education or environment. Today, the entire University of Washington administration is proudly behind the center, because we have broken traditional barriers of what academic institutions can do innovatively with published results. At the same time build demonstrable prototypes that require a lot of engineering excellence. We have done this by partnering with companies such as yours, for example, Zementis, and Microsoft to achieve social good. Apart from health analytics, we are working on genomic big data solutions for personalized cancer care. We are collaborating with various providers for predicting pregnancy costs for uninsured couples. We are one of the leaders in cyber security and developing cryptographic solutions for secure machine learning for (the) Internet of Things. We are analyzing teacher effectiveness pathways for special needs students with autism and blindness. We have a project going on that measures environmental impact and predicts which hazards will lead to the most amount of destruction or complications in our environment within the Puget Sound by integrating all this data, and it's hosted on Azure. We recently won (the) "Best Demo Paper" award for a prototype app that we developed on the cloud, which is called iTornado, that predicts… that uses complex predictive query processing for identifying safe shelters for drivers stuck in a tornado.

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878

MZ: Very good. That's why we feel honored that you actually chose us as one of your partners, really helping you with the deployment of predictive analytics in the cloud, for example.

When I think about specific use cases for predictive analytics, I often focus on four areas which we most often see at Zementis as client use cases:

1. Customer analytics: Customer behavior, looking at customer-centric data
2. Operational analytics: Machine-to-machine data, machine-to-machine communications; Today I think the hottest topic on that front is the Internet of Things and the ecosystems around that
3. Furthermore, for the financial industry: Risk Management Analytics, fraud detection, fraud prevention and risk scoring in general; And, of course….
4. and not to be underestimated at all, is Security Analytics: identifying vulnerabilities, cyber threats and similar issues that we face in an online world day to day, (whether) we see it or not

So, within these categories, what have you seen as some particularly compelling applications of predictive analytics? Who do you think is really leading the market in applying predictive analytics in these areas?

AT: So Mike, In my opinion, the data mining community has made tremendous contributions in all the above areas. In a purely academic opinion, and hopefully without offending anyone who might be a leader but I'm not aware of, I consider Netflix to be a powerhouse in intent analytics. They were one of the first ones to build large-scale recommendation systems and really contributed a lot to the academic community as well in innovation. Similarly, both Splunk and Bosch (Bosch is actually one of our sponsor partners at the Center for Data Science) in doing some very exciting things on the Internet of Things. For security, I have had the opportunity recently to collaborate on very exciting work coming from a new company in the space, called Niara.

Niara is elegantly using the big data stack along with scalable predictive modeling techniques to identify something that is a very hard problem in the security space. It's called insider threat analytics, and then the post-event forensics for example. So for these type of solutions, you need a highly scalable, very fast, retroactive vulnerability assessment solution that can consume large amounts of logged data, plus you need to be able to integrate intelligence across a variety of data sources for cyber security. So, Niara has been at the forefront of trying to design these machine learning solutions to do this analysis, and then it's an opportunity for data science in health care to not keep machine learning scientists awake at night! I think just the issues and the amount of opportunities in the cyber security space for machine learning are very motivating and should keep grad students up at night. So, that's the way I think about it.

MZ: Grad students could always be up at night … [laughter]

AT: OK, you say that - I didn't say that ….. [laughter]

MZ: If I may actually go back, when I started my question, I said there are 4 categories, but I just thought through your answer… I thought well, they are blending together. You have security analytics and the Internet of Things. Everything is online now, so security becomes even more important for the Internet of Things, and then likewise link into that customer analytics with your mobile phone / mobile device… it's a sensor but it's linked to you and your personal information. So now you already have three of those categories already blending together as one opportunity for data science.

But, let's turn to "the science of data science" for a moment: Do you find that there are certain predictive modeling techniques that really are more adept than others for addressing the unique analytic requirements in the healthcare industry?

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878

AT: In the healthcare space so far, the use of classification or regression techniques has found more prominence. Unsupervised learning techniques, clustering, have yet to bear fruit, mostly because in healthcare, the target variables are usually defined by the clinicians, or the physicians and data scientists are often invited to use classification or regression models to solve that specific problem so that clinicians can do a factor analysis of those solutions. We have been using naïve Bayes, SVM, random forest to solve a variety of these risk of re-admission problems. On the other hand, we have also been building some very large, scalable cost prediction models for looking at huge volumes of claims data and analyzing those claims in as real-time as possible, even though it's not really real-time, it's more batch, but as fast as possible in a scalable manner. And we have designed some very cool analytical models for predicting next year's cost of beneficiaries using soft computing techniques such as fuzzy rough sets or designing classifiers, particularly when there are a lot of inconsistencies in the data.

MZ: In terms of technology architecture, turning a little bit away from the algorithm to the underlying fundamental infrastructure, what benefits do you see from an enterprise-grade cloud environment such as Microsoft Azure, and how has a cloud environment been helpful to you in your work?

AT: We got our first Azure for Research grant from Microsoft Research Outreach Team more than a year ago. We recently also got a similar Azure Machine Learning Award to conduct experiments for readmission using Azure ML Tools. I feel that the life of a graduate student and faculty has been completely transformed by the advent of being able to do experimentation on the cloud.

Using SQL Azure and appropriately sized compute nodes, me and my grad students can slice and dice datasets much faster. We can share those datasets, issue large joint queries across multiple diagnostic scores and social demographic fact tables for our patients. And then we can augment the results with US Census Data and other tertiary data sources if needed and set up large machine learning workflows for building and testing hundreds of models in parallel. This was impossible to do before the cloud.

We can then cross-validate our models, we can select the best ones, export them to standards, such as Predictive Model Markup Language (PMML) or sometimes hard-code to PMML, and then set up scalable web services though APIs , such as ADAPA for scoring hundreds of patients or hundreds of patients' events in a very nice, robust manner. It also helps us test for failures and really try to understand where scale and missing values or attributes are not succeeding. So that is a plus of the cloud.

Moreover, since cloud services such as Azure are HIPAA-compliant, the healthcare providers are increasingly reassured that the machine learning models that we are building and the scoring that is happening is on a trustworthy, HIPAA-compliant cloud. That helps us accelerate our research using anonymized data without any PHI on the cloud.

MZ: Interesting. So, I guess it's fair to say, the cloud lowers the barrier for entry not just for the businesses to experiment, but also for research… for students. You don't have to wait for the next supercomputer cycle. You can actually go launch your own machine and run it. So it's really a fascinating topic and we could clearly go on for hours about this and still have much more to talk about. But, since our time is limited today, perhaps you could send us off with one last parting thought, kind of a next horizon. What do you see in healthcare analytics that you think is really exciting and will be the next big thing?

AT: So, it's not what I see. I think the question should be what do I NOT see as exciting or transformative in data science and healthcare. Everything from clinical integration of wearables to finding fraudulent transactions in claims data are in need of attention from data scientists. I would like to develop a personalized fetal growth curve for every baby in this world, particularly where malnourishment and HIV are causing havoc even before the baby is born. I would like to

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878

predict actionable interventions for neonatal care, identify learning disabilities in children through use of voice recognition systems such as Cortana or Siri or Echo, prevent abuse / domestic violence, promote wellness through wearables and safety. I would like to help depression… deal with depression through predictive gaming, for example, or reduce wait times for international adoptions by applying some very advanced recommendation techniques to match parents with potential children who are in need of good homes. Certainly I have done some work in the past on predicting calories from volume of food to help nutritionists deal with cancer and chemotherapy patients. In general, I think data science can transform healthcare by allowing healthcare professionals to partner with data scientists on improving protocols where ever protocols are amenable to data-driven decisioning. So, physicians will (one day, hopefully) also have to earn a degree in data science, or more data scientists will turn to become physicians and sort of transform healthcare one case study at a time.

MZ: It's fascinating to see that that's where all the exciting things happen. Thank you again for making this time for our conversation today. It has been a very valuable conversation for me, and I hope also for all of our viewers, whether they are data scientists, IT professionals or business decision makers, or if they work in the healthcare industry of course… physicians, clinical or otherwise.

I think to you, our viewers, let me leave you with one parting thought from someone who made a few of his own contributions to science during his career. Albert Einstein once said, "Look deep into nature, and then you will understand everything better." So, I hope he is going to be ok with us as data scientists that we say and modify his quote slightly to say: "Look deep into data, and then you will understand everything better".

[to Dr. Teredesai] So, thank you again for joining us today! [to viewers} Thank you for joining us! To learn more about predictive analytics and various practical use cases, please visit our website at www.zementis.com.

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878

About Zementis

Zementis, Inc. provides software solutions for predictive analytics. The company was founded on the principle that data science teams and IT departments can collaborate seamlessly and efficiently, allowing predictive models to rapidly move from development to deployment, so that businesses and other data-centric organizations can easily incorporate predictive analytics into their routine operations. Agile deployment of predictive solutions is the cornerstone of the Zementis philosophy.

CIO Review recognized Zementis as one of the "Top 20 most promising Big Data companies in 2013", and Gartner named Zementis a "Cool Vendor in Data Science" in 2014. Its ADAPA® and Universal PMML Plug-in (UPPI) scoring engines are designed from the ground up to benefit from open standards and to significantly shorten the time-to-market for predictive analytics in any industry. Customers such as Bosch, FICO, Equifax and Western Union have used Zementis solutions successfully to enhance their predictive analytics capacity and capabilities.

Zementis partners with leading analytics and data warehouse solution providers to enrich and extend customer capabilities. Supported partner solutions and platforms include: Amazon Web Services, Apache Software Foundation (Hadoop, Hive, Spark, Storm, Tomcat), Cloudera, Datameer, FICO, Hortonworks, IBM (BigInsights, PureData / Netezza, WebSphere), MapR, Microsoft Azure, Oracle WebLogic, Pivotal Greenplum, RedHat JBoss, SAP (HANA, Sybase IQ), Teradata and Teradata Aster.

For more information, please visit www.zementis.com.

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878