2018 NCSBN Annual Institute of Regulatory Excellence (IRE) Conference - Big Data & Machine Learning: What Does it Mean and How Do You Use it? Video Transcript

Event
2018 NCSBN Annual Institute of Regulatory Excellence (IRE) Conference

More info: https://www.ncsbn.org/11045.htm

Presenter
Ankur Teredesai, PhD, Chief Technology Officer, KenSci, Inc.
Whende Carroll, MSN, RN-BC, Director, Nursing Informatics, KenSci, Inc.

- [Ankur] So good morning everyone. It's a pleasure to be here. It's amazing. This is my first nursing conference. So I hope you guys are kind to me.

It is thankfully not my first talk on big data. So what I'm about to sort of show you today and explain is a little bit of personal journey through the history of medicine for me.

I'm a computer scientist by trade. I grew up learning about programming and algorithms and pattern recognition and trying to solve problems using a box, right? And then I stepped out in the real world and tried to understand how this box could be useful to people.

And that's where the real challenge began. So the agenda that we thought might be great for conversation would be to just go over and explain some basic concepts of big data and try to clear the air around what big data means, why we are investing heavily in machine learning, why is the world going crazy with AI, and how Death Robots are going to take over the world and manage, you know, everything in human life.

With nurse Wendy, I collaborated for over a year now on working on various health scenarios that affect nurses directly and she is going to present a case study of the work that we're doing at EvergreenHealth and that will give you some insight into how big data principles that I will explain actually get implemented in the care setting.

So hopefully, you know, we'll get a lot of good conversation and learning from that. And then, I will outline some of the challenges and opportunities of collaborating together as some of the world's best leaders in this field are here in the room today, it is my pleasure and honor to be able to share this and discuss and get feedback from you whether, you know, the world of AI and machine learning is thinking in the right direction when it comes to, you know, forming this tight integrated learning loop between both of us.

So I hope that this is as much as a conversation as it is an educational component to this program. So without further ado, let me go into the details of big data and machine learning. So the first question to ask is why do we care about big data? Right? So any takers? Why do you think we should care about big data?

Or assume you don't know anything about big data and still you're asked to give an opinion on why you should care about it, that's the best way to start a talk, right? But I'm sure you've an opinion or you've read about it or you think about it. So maybe one or two volunteers would be great to start.

- [Woman] I think it can improve efficiency and reduce error rates at least in nursing. You know, we have so much of it and we collect it, but it's good to know what it means so we can apply it to reduce our workload.

- Very nice. So improve efficiency is the biggest draw for trying to do big data. Any other opinions? - [Woman 2] I think big data also may allow us to make more credible generalizations and I think big data can also show us some trends that we then may need to dig into with smaller data to understand the why that gives us a big picture of what's happening.

- So efficiency and precision, right? Efficiency and precision are two things that come out very quickly when we think about big data. But what if I said those are the wrong reasons? What if I said the fundamental reason why we should look at big data is because we're a part of an industry that thrives on one single world called empathy.

What if I told you that the fundamental principle of big data should be to promote empathy? Would that change your perspective about big data? Right? Think about it. The reason I love data science and machine learning is because I feel that we've enough horsepower now in our computing world to actually focus on some of the bigger problems that we deal with today.

Empathy and humility being one of the biggest challenges that we can teach computers to think about. The big challenges that the world of big data is solving is not precision or accuracy anymore. I can give you tens and thousands of examples of Siri, of Alexa, of many, many other tools that are very accurate.

It's a matter of time before that accuracy can be included to a great extent in the world of medicine. Not to replace what you do or what the physicians do in their practice, but to assist and bring in that sense of empathy again and again and again in your practice.

And I'm going to show you today how we do that. So why big data not for precision, not for accuracy, not for efficiency, but empathy? What does it mean for our patients? How do you interact just because you know you have more data about a patient? How do you use it at the right time, at the point of care, to make it more actionable?

These are the questions that the healthcare artificial intelligence community is asking itself. How do we use it? Does it impact me? Is it going to change regulation fundamentally? Are we going to be suddenly more intelligent because we have more data or we have better data?

What are the practices and behavior changes that need to evolve through the regulatory bodies, through policy that we need to discuss in order to make that data appropriate and actionable, are some of the questions we need to ask. Of course, we need to ask is this all hype?

Are we ever going to get there? Is this real? You have that question, I do too. And we'll show you that some of it is hyped, but some of it is real. So the answer is a mixed bag, even today. And probably will be for the next 10 years, at least. The last point that I will emphasize and discuss is are we ready?

Because the worst disservice that we can do to our profession is to promote something that we're not ready for. Right? If there is something to be said about timing, there is a reason why we developed ICD codes, right?

And then they got misused all over the place, right? And there is a reason why Florence Nightingale decided that she needs some form of a coding system to keep track of what's happening to patients at a large population level and that whole thing evolved into a coding system and now hundreds and thousands of companies all over the world and entire 3 to 10 trillion dollar industry across the world is basing its foundation on these coding systems.

What does that mean? How do you evolve it? And how do you get started on this journey of big data? By a show of hands, can you raise your hand if you think that whatever I have said so far makes sense? Okay, thank you. Just wanted to make sure I'm in the right concept. So here is a little tidbit about me.

I'm actually a professor of computer science and system at the University of Washington. I started my journey in healthcare around 10 years ago when I was first presented with the problem of trying to identify patterns in Oncology and Nutrition at the Frederick Cancer Research Center.

Folks came up to me and said, "Hey, we sent home a lot of chemotherapy patients and we really don't understand what's happening to their body because we're in the era of precision medicine. So we no longer want to know why someone has cancer, but we also want to know why someone has cancer of this particular type, of this particular nature, of this particular characteristics.

So we're going down the precision chain. And in order to address that problem, the parallel in the human behavior world was nutrition. As cancers get specific, so does nutrition. We all eat very different things, right? And that was a very interesting formulation to understand. So I got involved in that study. It was a very large NIH grant, multi-year and we solved a lot of interesting problems in it.

We even designed a very interesting app on a phone to take pictures of food, to take videos of food, and try to understand the calorie counts of how much people are consuming at that time. This was 10 years ago. I worked previously at Microsoft and IBM. So I know a little bit about how enterprises think when they try to solve large problems of software development and together all of that is what I'm bringing to founding the company called KenSci.

It's something that started, you know, two and a half years ago and I will tell you a little more about that. But before we do that, I'm on one mission at KenSci and that mission is to fight death with data science. Okay? Death with data science. The reason for such a bold mission is to again not design solutions that make mortality easier or palatable or empathetic, but rather to find solutions that help that journey of disease progression and risk management in the entire system.

I do that using two insights: Detect variation and do prediction. How many of you think variation in your daily lives? A lot of you. You may not realize it. Every time you see a patient and you think about their unique characteristics, you're doing something different to that patient, compared to a similar looking patient in your ward at the same time.

The reason for that is inherently we think we personalize things to a point where that variation sometimes becomes so large that it's very hard to control. And variation analysis is one of the fundamental tenets of machine learning that we see greatly being helpful in healthcare. Of course, all of you know about precision medicine.

You read about it. And precision medicine is more towards prediction medicine where you're trying to pose a problem, will this patient be diabetic in 5 years? That's a prediction problem. It's not a variation problem. A variation problem is, amongst the group of 10 physicians in the health system, adjusted for risk outcomes of all the patients that they treat, what is the variation between these physicians?

That's a variation problem. These are two fundamentally different problems. If we solve for variation and prediction, then we can solve for right care, right patient at the right time. That is the fundamental thesis of big data. Going higher in that hierarchy, you can start seeing segments of solutions emerge along cost, around provider experience, around patient experience and engagement, and care management.

So it's a network effect of solving things. So I'm boiling it down and I'm going bottom up, so that you all understand the different parts where big data and machine learning are trying to change the game. And finally driving all of this innovation in the world today are the volume of data, the variety of data, the velocity with which data comes to systems.

The veracity, the truthfulness of that data at source and after it goes through some system, and the value that we expect out of that data. So this landscape of thinking is what you should take home with you to see and understand and reflect how big data is changing the world of healthcare.

So we all start with the mission and each of you may have your own different mission, but the template is very similar. If you boil it down, it will come to detecting variation or targeting some prediction problem and then you can move on up. Any questions? So here is KenSci, I said our mission is death versus data science. We're a software technology company but we've physicians and nurses and program managers who work closely with us.

We don't do things in isolation because very early on, we learnt that software for the sake of software gives you a clippy. You know what that joke means? A clippy. Remember in 1996 Microsoft came up with a clippy that as you started typing in Word, it would automatically tell you "Oh, I think you're trying to write a letter, let me help you with that letter."

The idea is to not go and make that same mistake again of developing a clippy, so we want the tools that we create to be accessible, to be usable, and so we work collaboratively with a large body of domain experts. We're quite well funded and our headquarters are in Seattle with offices in Hyderabad, Singapore, and we work also with the NHS in UK.

So we started with machine learning research in my group in the University. We teamed up with Microsoft for a lot of our technology partnerships and then we brought in the healthcare expertise to guide us. Together we are KenSci, but machine learning is everywhere, you know that already.

Here is an early illustration of what was expected from machine learning. The idea was that you give...how many of you have heard of the Turing test? Some of you? Okay. Alan Turing was a very famous mathematician. He was the one who helped break the enigma code in World War II and he came up with this test to say if you can design a computer that can take any random input and give an output, right, and if there was a human sitting at the other end who was listening to that output or reading that output and if the human could not decide whether this output was generated by a computer or by another human, then that computer has passed the Turing test.

So the gold standard of AI and machine learning is to pass this Turing test, right? And think of it. Alexa is coming pretty close to that. Well, not generalized, do not generalize AI, but at least for question-and-answering. You can also see glimpses of attempts to make Turing test accessible or to beat the Turing test in many domains.

I've a colleague, Oren Etzioni who works at the...he has an initiative called AI square where he's trying to get computers to solve eighth grade math homework. And there is billions of dollars of funding in trying to find computers that can solve eighth grade math and reading exam.

If they can pass that we think we have achieved a lot, right? But we are not there yet. AI is not there yet. But it is there. It is getting there. For example, email, right, whether email is spam or not today we don't have to think too much about spasm because somebody solved that problem using AI for us.

You know, whether, which products we would like to buy next on Amazon, it's a pretty well-defined nail down problem of recommendation systems and we can solve that problem. So in certain places like face recognition. You know, China has implemented technology to scan faces on the street in any major Chinese city and automatically identify every single individual that's walking on the street.

That's possible, right? And, of course, we do fun things with AI, try to identify cats and dogs. But going one step deeper, I was asked by Noor when he invited me for this presentation to make artificial intelligence and machine learning intuitive and understandable to everybody in the audience.

So I'm going to use a little bit of math, but very simple math to help you understand how does machine learning work. What does it do, right? I hope all of you are curious as to what it does under the hood. So let's look at an example here. These are data points on a block. And there are two types of data.

Take any example. Diabetics and those who are not diabetics. Chronic patients versus non-chronic patients. Patients who will readmit versus patients who will not readmit. You can pose this problem in many different ways. The greens and the red and the blues are well separated for most of the space on this block, but there are regions where there is a lot of confusion.

And even a human eye cannot correctly ascertain which one is green or red because they're overlapping.

So far so good. Great. Traditional statistics, and you must wonder why is machine learning so different?

What's it about big data? Come on, people have done math and statistics for centuries, why are we talking about this now? This is what traditional statistics would do. It would come up with an equation, a mathematical formulation that would create a linear line between the best possible separation of these two classes.

Our idea is to come up with a computer program that can help if a new point was put on this plot to help identify if that new point was green or blue. That's the problem we are trying to solve. So we have all this historical data and we're going to get a new patient and we don't know where...we know where the new patient is going to be on that, but we just don't know what label to give to the new patient.

So here is a classifier that simply takes of a measured judgment and says, "Ah, because majority of the green points are above this line and majority of the blue points are below this line, I'm going to have a linear boundary that separates it. What's the problem with this? The problem is the error and we talked about medication errors.

We talked about efficiency. The problem is when you use a solution like this, there is significant confusion that happens in the bottom region. So if a patient who looked like they're all diabetic patients was to come and if that point was, right down there at the bottom somewhere, this previous statistical technique will not be able to solve that problem.

There are regions of confusion that don't work well with this technique. What machine learning does on the other hand, is it does away with this one simple line and it tries to estimate more precisely, the curve or the boundary that separates majority of the blue labels from the green labels.

Of course, it requires more computing, right? It requires more hardware, let's put it that way. When it requires more hardware that means it needs more time and more effort to learn that pattern. But with that time and resource and compute power, today it is possible to come up with a more precise boundary definition that solve the problem more accurately, more precisely.

Does that help? So that is the key difference between traditional statistics and new-age machine learning. And why did this happen? Why are we able to do this? Why big data? Anyone? Why couldn't we do this before?

That's the question. We didn't have the compute power, that's right. That's one part of it. We didn't have the compute power. So computing technology has advanced to the point where it is possible to do this in the same amount of time that it was possible to do that previously.

We carry phones that have more power in our pocket than the entire Apollo mission. So we are there. That's one reason. What is the second reason why we need technology like this and the answer is right there on the screen. Complexity.

This is a two-dimensional plot. Imagine if each data point there was hundreds of dimensions going all out of that screen in all different directions. Imagine yourself in a 3-dimensional, 10-dimensional space where you have this point, and now you're trying to solve the complex labeling problem across hundreds of dimensions, right, it becomes intractable to solve very quickly.

An example of that, a real-world example is a patient 50 years ago and a patient today. So think of a patient pre-EMR and think of a patient post-EMR. Pre-EMR on a chart we had 20 things that we knew about a patient.

With the EMRs and with digitization, now we know 500 things about the patient. So a patient that was a unique record with only 10 attributes just 20 years ago is now a patient with 500 attributes, 500 different dimensions to represent that patient.

To handle this complexity, if we use the traditional statistical technique that could handle only 10 attributes, it doesn't give us any better results. So all that new data and investments in infrastructure and technology that healthcare has made in the last 20 years goes to waste because the majority of time is spent in bringing that big data down to small data and then doing statistics on top of that data.

So you lose a lot of information and you use a lot of fidelity and you lose a lot of compute power in just trying to reduce the number of dimensions along that patient. Now, imagine if you could do machine learning and AI without having to reduce those dimensions, suddenly, it would be much more powerful, it would be much more effective.

Makes sense? What I'm showing here is a classic decision tree algorithm. It's like playing 20 questions. I ask you to guess a name of a place or an animal and then you ask me the right question and if you can solve the problem in 20 questions, you win the game. If you are not able to solve the problem in 20 questions, I win the game.

So the challenge is to teach a computer to learn which is the best question to ask at the beginning. And slowly when we start playing this game with our children, we quickly learn if it's a place we start learning that it's best to ask whether it's in the northern hemisphere or the southern hemisphere, whether it's in the eastern hemisphere or the western hemisphere, things like that.

So we're teaching a computer to ask the best question first. And that's the name of the machine learning approach. So what types of machine learning? How many types of machine learning are there? Well, two main types: Supervised learning and unsupervised learning. Supervised is when you actually have labeled data.

Supervised learning is like teaching a child, this is a book. This is also a book. This is also a book. This is a book, but it is a fat book. Right? So this is a book, it's a children's book. So slowly the child starts understanding, book.

The next time they see the object book they know it's a book. This type of learning is called supervised learned. We are teaching by examples. There is another type of learning, which is unsupervised learning. Child trek or hike lake leave the child.

Nobody does that, right? But in machine learning there is utility to doing that kind of stuff. You say, "Hey, we don't know what the goal is, but somehow try to in an unsupervised way, find out groups, patterns, clusters of data that then makes sense, and come back to me with an answer and I will tell you if the answer is a good answer or a bad answer."

It's like, John, go fetch a book and John comes back with a pencil box. "Not a book. Go fetch a book." Then John comes back with a bagel. "Not a book. Go fetch a book." So slowly through the repetitive process, John learns that that is like a book and this is like a bagel.

Makes sense? So you're not supervising, you're just validating whether it is good or bad. That type of learning is called unsupervised learning. Most of the recommendation systems that Amazon uses, for example, is unsupervised. Why? Because, you know, like Netflix. How many of you use Netflix?

Everyone, come on, raise your hands. That's right. Netflix learns not by asking you. Well, they only ask you three movies that you like, they need that. But after that they don't keep asking you whether you like this movie or you didn't like this movie. They just learn in an unsupervised way, based on patterns of your viewing behavior with hundreds and millions of people across the world.

So I might be more similar to Linda because we both like a particular movie and then as the distance between what we like decreases, it starts recommending things that she has viewed, but I have not viewed before. So it starts calculating that I'm very similar to her in terms of watching behavior or what I like to watch.

So in KenSci, the types of machine learning we use is supervised unsupervised, but we also solve problems through that space and we've solutions for risk of readmission prediction, end of life prediction. We're just presenting a paper next week at the AAAI conference, which is the flagship AI conference in the world on predicting 6 to 12 month mortality prediction with around 80% precision.

So that we can help physicians and nurse practitioners have a better conversation with the patients on palliative care, on reducing their footprint within the acute care setting and so on. So we've done a lot of work on cost prediction, for example. We worked closely with the CDC on predicting future healthcare costs of 9/11 victim families and first responders because under the Zadroga Act, CDC has been mandated to underwrite and insure 9/11 victim families and first responders and the work of predicting how much that population is going to cost over the next 75 years is the work that we are doing in partnership at KenSci.

Just an amazingly interesting project. Of course, we do a lot of work on ED load prediction and we're going to talk about that sooner. So the steps of machine learning, how do you get there, what do you do? Well, you get the data and data is very important. So we need clean data, but we never get clean data.

So we need to invest our effort in cleaning the data, preparing the data, shaping the data, and then we train the machine learning model. We train the program to produce a program that will then give the output when a new data point comes in. Then we test it and we improve. So it's a repetitive process. So from data to machine learning models to two types of models or insights, variation insight or prediction insight, and then that leads to actions.

So I will stop here. I'd welcome nurse Wendy to come up and present our efforts with EvergreenHealth and talk a little bit about her experience at KenSci. So welcome Wendy. - [Wendy] Thank you. So big data.

There was a time when I knew nothing about it at all, I was completely overwhelmed by it and I was a complete deer in the headlights. My personal journey with big data started when I was a population health management data scientist at Worksite Wellness Clinic, and I didn't know and I was new in that role, I was going to have to aggregate and analyze so much data from so many different sources, I was absolutely again just terrified.

But out of being terrified, and having to learn it, I became really interested and curious in what it could do and like what Ankur talked about how it can bring us to a place of empathy, particularly in population health management. So that's one of the ways that I used it and that's my story.

We're going to talk today about a community hospital in the Seattle area called EvergreenHealth and how we helped them to predictably match their ED staffing with demand.

And the reason why they wanted to know this and it was so important for them to know this was because they were having a lot of staffing problems in their ED. They were having a lot of issues with patient satisfaction and nurse satisfaction and that was causing a lot of challenges along with that that was costing the organization quite a bit of money.

So what they wanted to know from us was can we predictably use our data to match nurse staffing with patient demand. And we knew that we could help them but we knew that it wouldn't be without challenge. The data currently shows that ED demand for services is predicted by weekly and seasonal patterns. We already knew that.

That was a good news, that prediction was possible. The not so good news was that we found this prediction could be difficult because we had to control for the arrival rate of the emergency department patients coming in and because there were significant pattern differences and we would have to control for temporal climatic patient factors that influenced that particular arrival rate.

And if you looked at staffing, staffing was even more difficult to predict. If you look at all of the arrivals from the year 2014 and 2015 trying to predict for 2016, of the arrivals that were coming into the emergency department, you'll see that there is a lot of variation.

And you'll see if you look at staffing for the same time, that staffing stays static regardless of the variation and the drastic variability that they have in their arrivals into the emergency department. And the reason that is, is because staffing is done weeks out. It's used based off the rules based systems that are heuristic, guessing methods with little or no structure data at all.

So we also found in their data, when we're looking through it that they had a lot of something they didn't really realize was from a cost factor a lot of overtime hours that were being used. If you see on the big bubble heat map here, you can see at the top there are some outstanding red dots and those are overtime hours and those are high overtime hours.

And you can see that it makes sense from a nurse-to-patient staff ratio perspective when there is more patient-to-nurse staff ratio that should be higher. But if you look at the patterns we detected within the bubbles, you'll see that there was a lot of overtime hours being used irrespective of if there was a nurse-to-patient staff ratio or there was a time of low census. So that's something from a cost perspective that we really found that was valuable to them to be able to predict as well.

So to help with their variability and with their decision making, we developed a couple of few kinds of solutions to be able to help them tactically, and then also from a strategic standpoint. So the first thing we did with our machine learning platform at KenSci when we went to Evergreen was to look at their situation and what we found looking at their data was that they were having a lot of overcrowding that was leading to a lot of patient dissatisfaction and nurse dissatisfaction and again the cost associated with that.

They were also having that overtime usage that we found as well. So the tactic was that we took their EHR data and then we made predictions for two, four, and eight hours prior to patient arrival. We looked at the predictions down that far.

And then we were able to give weekly and monthly demand pattern predictions to help with the staffing requirements and then help that ED adjust their leaderships staffing as well as look at their tactical real-time at the point of care staffing.

So the first thing we did was to take a data as Ankur talked about unclean data noise to a useful, meaningful signal was to look at their historic data for two years and we took a look at it by arrivals per shift per month.

And then we looked at another longer term pattern to show their historic data and then what it would look like real-time and then what the predictions would look like. We took that prediction down to one hour pre-arrival. And that's that tactical response we talked about and we'll look at another piece of that in a second.

So what we found which we're really excited about was our predictions were statistically much better than their prior baseline methods that they were using with their heuristic methods and their rules-based systems and their guessing and guesstimations with no or little unstructured data, compared to what they were.

So we're really excited to find that out. And again to move from that noise to a useable, meaningful actionable signal, we developed a long-term forecast for a six-month long strategic planning, so nurses and administrators could see that overcrowding and also ensure that there was better nurse-to-patient staff ratios.

And then a short-term real-time dynamic visual that would help to be able to help charge nurses control that overtime usage in the EDs at the point of care. So that's what we did with machine learning for Evergreen to directly be able to impact a nursing workforce issue that they were having.

We really focused broadly, looking at it from an empathetic level, looking at it from the quadruple aim and particularly lowering costs, which is one of the things we're able to do and then as we talked about before, lower waste, higher efficiencies, more productivity, better nurse-to-patient staff ratios, less staff burnout, so that staff overall improvement and staff satisfaction and then ensuring also as we talked about at the beginning as well that we've lower turnover and higher retention and lower cost played into that as well.

So coming from an empathetic angle and looking at it from the quadruple aim, that's for we are really starting when we looked at EvergreenHealth. They were really happy with the...as you can see here, they found exciting ways to find pattern in massive data that lead to actionable change and can tangibly impact patients and their outcomes and also the outcomes and the satisfaction of nurses.

So Ankur is going to continue to talk. Thanks, Ankur.

- Thank you. So that gives you a glimpse, just a glimpse, and there are hundreds of problems like this to be solved. One of the first challenges that we ran into at Evergreen was, you know, they couldn't even give us accurate data on how many nurses are in the ED at a given time.

They couldn't. There was no system that captured that accurately other than the staffing roster. So we had to go by the schedule, but you know that schedules change by the hour.

People don't show up, they have emergencies, stuff happens. So there is no unified system that could correctly record what is the right nurses-to-patient ratio at any given point of time. So real-time is a problem, but even long-term is a problem, especially if not ED, at least, in the OB setting, right?

How do you know what other staffing levels needed? Now, Evergreen is in a very interesting neighborhood in Redmond and Kirkland, where there is significant growth with Microsoft and Amazon in Seattle, that region is booming. So keeping up with that demand and just childbirths is a huge problem for them.

And some of their national metrics on which quality of childbirth wards is determined is really suffering because of that growth. And that's a real challenge that they're facing today. But for our conversation, we thought that we would propose some ways in which NCSBN and policy leaders here in the room could think about machine learning and AI in their next year's planning, let's say.

So one would be a variation problem. Again, I mentioned that there are two types of problems: Variation and prediction. So the first problem is really how do you determine or control for variation in the quality of nursing applicants? How do we do that?

Can data help? Can big data help? Can machine learning predictively solve the problem of this applicant variation? Any thoughts? Is this a problem or not a problem at all? Yes, please.

- [Kathy] Sorry, I can't help myself. I'm Kathy Giblan from Alberta, Canada and I'm going to be speaking after lunch. So I'm not going to give away what I'll be talking about, but I will answer you yes, this is a problem, problem my organization faced directly with huge volumes, huge numbers of applicants as a result of overseas recruitment efforts by employers in our context and lots of pressure to make it happen faster and lots of pressure from employers who should we be targeting, who are the good ones, who's going to get a license and get one quickly and lots of variation in the characteristics of our applicants, like huge variation.

So the answer to that question is yes, and we wanted to take that variation and actually make predictions.

- [Ankur] Correct. That is the next...Yes? Thank you so much. -

[Woman 2] I still am kind of on the edge of really believing in big data because I take care of patients in a rural health clinic and I get a lot of patients that come from big data EMRs, but if they're not intra-operatable, you're not going to get...you are going to miss a lot of things and I'm actually seeing some of these big, like e-clinical works and things like that. There have been some miss-steps where if you don't talk to the patient, you just rely on the data, you're missing, and you can make some really amazing miss-steps in taking care of the patient.

You can lose a life if you just look at the data and don't look at the patient. So I think that, you know, it's as good as the person I'm putting it in. It's not infallible. - [Ankur] I completely agree with you. You're stealing my punch line, but you know, the last slide that we have is my passion is to reword the word or the phrase artificial intelligence to what I call as assistive intelligence and many times when we forget that machine learning or AI is not by nature assistive, we run into these exact types of problems.

So just by having data or having wrong data can cause a lot of problems. If the underlying technology that is parsing that data, looking at the data, and supporting clinical decisions is not assistive by nature, so absolutely right. There are lots of gaps in the system today.

We're not perfect, but this is one way of getting there, right, and the reason that it is important to talk about that in this context is because by understanding variation within the quality of applicants or within the capability of applicants that apply to become a practicing professional, you are starting at the bottom.

You're starting at the source of that behavior, right? So this is just one prong in a many pronged approach that needs to be taken across the system. So while I completely agree with you, I'm also trying to explain that there is a place for where there are gaps today and what is happening is we are expecting magic to come out of downstream systems when fundamental systems of training, education, and regulation are also broken.

So that's the main point that I want to drive towards. So moving on to the next one, the next picture. Sorry, yes? Please. Yes? Let me paraphrase that so that others can also...so may I have your name? -

[Leah] Yes, Leah Philipps.

- Sorry?

- Leah Philipps.

- Leah Philipps.

- So Leah Phillips' point is very well placed. She's raising a question about the volume of data. And this comes up in machine learning and AI again and again. The reason, have you ever thought why Facebook is able to tag you the moment you upload your picture or someone else uploads your picture? The reason is, it has millions and billions of images that it has processed.

So one of the fundamental things in big data is actually big data. So if you try to solve that same problem using small data which is 60,000 applicants in this particular case, you may not find that great of a variation. But even if you start with 60,000 applicants, and looking at longitudinal history of those applicants or looking at secondary data sources or forward behavior of those applicants as they go through the system, suddenly the volume of that data start becoming bigger and bigger.

So you have to start somewhere, right? And I always recommend starting with small data. Because once you start with small data, you're putting the instrumentation in place to start collecting that big data and managing that big data. But if you shy away from the promise of small data and small variable statistics, then you will never make that leap of faith and by the time the problem becomes chronic, it becomes unsolvable or intractable.

So I encourage the efforts that you're taking in working with small data and I encourage you to think of how you can join your data with other data sources to make that data bigger data. So that's one solution to the problem. So there are other prediction problems, you know, nurses who might be propensity of being fraudulent.

There was a nice article in the Journal of Nursing Regulation that appeared just this year that talked about the use of big data. I think it was written by you, Mary. So I want to give you a round of applause for writing that. Thank you so much. And then respondents likely repeat serious errors.

Is a coding error or prescription error or just management error something that people deal with on a daily basis, but can you predict that ahead of time? Are there behaviors you can see ahead of time through other smaller errors that then amplify themselves downstream in the system?

And how do you put actionable education behind it to help catch it at the right time so that you can focus on making it better. And then the last one is really evidence-based predictions. So, you know, is there really a holy grail of using big data in this situation would be to have licensure based on multiple variables from several data sources, you know, profiles and so on.

Now, again, there is a fear of what does that mean? Does that mean you know, before giving a license we should look at someone's Facebook profile or look at their political views or look at other things, the purchase history of what they buy in superstores? No, that's not the idea here, just to make sure that I'm reassuring you that that's not where the world is going and I certainly don't want the world to go there, but there is a bounding box that is so small today.

If it was expanded a little bit with the right constraints and regulation and policy, I think we can achieve a much better outcome. That's the objective of the conversation. So how do you apply machine learning in healthcare?

There are four important steps. First is to identify a problem, an important problem. How do you identify an important problem? I measure importance by only two metrics. Either there's a huge return on investment from a financial, from impact perspective. Or, the problem has such societal impact that it will fundamentally change the way we act in medicine.

Those are the only two pivots for me, which are important. And if we can qualify every problem that we are going to solve in healthcare using machine learning across these two pivots, then we're on a success track, and we're doing that. Today we're doing that for many health systems. Then, the next thing is to get the buy-in from the stakeholders on the data driven approach.

We all, at least for me and Wendy, we interact with health systems all over the world where the CIO or the CTO, or someone in technology is interested, right?

The CMO and the CNO are very brought in, right? But the downstream enablers are not brought in to the change at all. So at the C level, you'll get a lot of buy-in. But if you don't have that buy-in at the troop level who is actually going to use the outcome of this tool, it becomes extremely hard to implement the machine learning approach, the data driven approach.

So there's a fundamental need to address that problem before you start or you embark on a journey through data driven decisions, right? Not that C level suite doesn't make the decision, but it is very important to find champions at each level of the hierarchy in an organization that will then enable data scientists and machine learning people like me to come in and change things.

So change doesn't happen at the top. Change actually happens at the bottom. So we got to influence, educate, discuss, talk, communicate, and share the benefits of what we are going to do and get that buy-in ground up before we start embarking on machine learning projects top-down and that's been my key lesson as we work through this in AI and healthcare.

Then we need to move from variation to prediction. I think we all discuss that significantly here today, so I won't go into the detail. But one more thing that we have found to be successful in healthcare AI is go for a quick win. Show that there is small value, incremental value, right? So pick a project within your organizations that will lead to a quick win.

If you go for a very long-range project that requires investment of multiple years to get an outcome, then you will fall into this hype cycle and then nobody will believe you. So we always like to qualify the problem to make sure that it meets all these four attributes in your system, right? So try to find problems and then think of this framework.

Does it meet that? Does it meet that? Does it meet that? Does it meet that? If it does, you probably have a good problem on your hand that someone can solve using data science. So healthcare machine learning is hard and even today, healthcare is still the hardest part, right? It's not the machine learning anymore.

I have hundreds of students who have graduated under my tutelage at the University of Washington, they are really good mathematicians, there really good computer scientists. The math is not the problem. Healthcare is the problem and we need to fix healthcare, not the math. So are you ready for machine learning in healthcare?

Here's a quick way to evaluate. Just look for data-driven decision champions. Look for data readiness. Think about governance structures if they're in place or not because guess what, you may not like the outcome that you see from an AI-driven approach or a data-driven approach. Do you have the right governance structure in place to ensure that the patient outcome that you're driving for are controlled property?

Is the security, is the privacy of those patients being taken care of, once you find out the insight that you're looking for in that ecosystem? All of these things need to be thought about before we embark on AI machine-learning project. Are your stakeholders convinced? Are the policymakers and regulators on board with this?

Because this is a tsunami, it is coming. If we're not prepared for that tsunami in the right way and we don't have right regulation in place, then it will become too late. And if we try to retrofit regulation to the solution, it will become very challenging later on, because guess what, Facebook would have already been invented and, you know, fake news will already be there and now if you try to retrofit tooling to ensure that fake news is not propagated on a platform like Facebook, guess what, you're putting the cart before the horse.

So that's why it's important for you in the room, especially, who are at the forefront of this thought process to be thinking ahead whether you have the right stakeholders who are invested and convinced on this or not. And there are, you know, many machine-learning vendors, many platforms, many decisions to be made on technology, and I'll save you the trouble of thinking about all of that for now .

Again, my ending message would be to start with a quick win. Look at the data, try to seek information out of it, that will lead you to some knowledge, that knowledge leads to some insight. Make that insight actionable and once you make that insight actionable, you will find wisdom through this path.

And continuously learn. So a machine-learning system is good when it is learning, getting feedback all the time, and learning continuously. So design a system that goes through each of these steps and is still continuously learning because it's not going to be perfect. I like to quote Jürgen who is actually the father of all the convolution neural net technology that is now powering Alexa and Siri and deep neural networks today and his famous quote is "We are at the beginning of the end of the beginning of AI."

And I leave you with that philosophical thought. Thank you so much for the opportunity. I appreciate it.