Stanford eCorner

Shaping the Future of Data [Entire Talk]

Gurjeet Singh, *Ayasdi*; Gunnar Carlsson, *Ayasdi*

February 05, 2014

Video URL: http://ecorner.stanford.edu/videos/3266/Shaping-the-Future-of-Data-Entire-Talk

Ayasdi Co-Founders Gunnar Carlsson and Gurjeet Singh discuss building a company based on technology developed inside a university, in conversation with Floodgate Partner Ann Miura-Ko. With deep insights on data visualization, Carlsson and Singh talk about developing technologies and the process for evaluating high-impact applications.

Transcript

It's a real pleasure to be introducing Ayasdi here and the co-founders. Just briefly, I'm going to give you the titles of the gentlemen sitting to my right here. This is Gunnar Carlsson, who is also a professor here at Stanford. He holds the Anne and Bill Swindell Professorship in the school of Humanities and Sciences. And this is also Gurjeet Singh, who is the CEO of Ayasdi. To tell you a quick story about Ayasdi, I met Gurjeet back in 2010 and ... 2009. ...2009 and he had... a friend of ours, a mutual friend of ours had sent me four math papers that Gurjeet had written, asking 'is there a business here'? And I had just started off my career as a venture capitalist at this point and I was also simultaneously trying to finish my Ph.D. And so I was in this mode of reading papers, and started going through Gurjeet's papers and realized that there were some really interesting insights that he had, and I felt like there was a business there.

I then met with Gurjeet and became further enamored with the business, and with him as a founder of the business. And then went on to meet Gunnar, and at that moment I just felt like this was the business that I had to invest in, that no one else was allowed to be a part of this business at the seed stage. And so, I managed to chase Gurjeet down to a classroom at Stanford and then offered him a check for $1 million. And he was disciplined enough to say no, because he didn't really know what he was going to do with the money that I would give him. And we proceeded to work together for a little while. He would do some customer development and come back and make me even more tantalized with the business. And eventually I wore him down to the point where he would take an investment from me. And since then I've been so glad to be a part of investing in this company, seeing it grow. But the story here is really one of some real amazing technical insights coming to fruition in the form of a company. And especially within this audience, of people who have background in math, engineering and sciences, this story is very much relevant to you.

And so I hope that speaking to Gunnar and Gurjeet today you will get a flavor for what it takes to take some real technical insight and turn it into a company. So to get started, I would love for you guys to tell us about what is Ayasdi? Awesome. So we're, by the way, thank you for inviting us. It's a real pressure to be here, speaking with everyone. I remember sitting in the audience for pretty much two semesters, that was incredible. And I'm just so happy to be, to be able to speak with everyone. Ayasdi, so we're trying to solve a pretty big problem. We are trying to solve the problem of turning data into knowledge. So the standard way in which you turn data into knowledge is by this process. So you have a smart analyst, they come up with an idea or a hypothesis, and then they might convert this hypothesis into some sort of query, some sort of a code.

Or they might use some business intelligence software that converts it into a query or a code. You run it against the database and you see the results and maybe you're correct, maybe you're not. And that's the standard process by which we turn large amounts of data into knowledge. And you know there are a few problems with this process in general. The first problem is people. So usually you need really specialized people who can come up with these hypotheses, they are really complex. If you look at the search volume for the term 'data scientist', this is off what the Google trends will tell you, in 2011 no one knew what it was, in 2013 everyone wants 10. And the other thing about being a data scientist is that you can't drop out of school to become one, you have to have some sort of an advanced degree, you need to know some combination of mathematics, some statistics and some computer science and some domain information to be able to become a data scientist. And the second problem with this picture is hypotheses. And if you think about just a very simple tabular setting, hypotheses are essentially a subpart of a large table.

And if you think about how many hypotheses exists in a table, it's usually exponential in the number of, and the size of the table. So there are too many hypotheses, and that's a big problem. What we have developed at Ayasdi and what our point of view, is to use a much more automated methodology. So we basically take large amounts of data, pump it through many many hundreds of machine learning algorithms. We are able to combine the results of these algorithms based on our research back at Stanford, that we might discuss during the, during this talk. And the first time a person enters the picture they already have some answers to begin with. So that's the basic idea. Our business is about turning data into knowledge automatically in close to zero time. Does that make sense? Great. Well, Gunnar, I know that this was a result of a few decades worth of research for you that finally came to fruition here within Ayasdi.

Can you give us a little bit of that history? Sure. So I started out... I'm a mathematician, and for most of my career I've been a pure mathematician. That is to say in particular in the area of algebraic topology. I did my Ph.D. here at Stanford, and worked in that area. Which is, it's a very old part of the subject, it's also gotten very esoteric to the point where people don't, people outside the small area don't always understand what it is about. But from my point of view there was always this idea that we should keep our eyes open for opportunities for using some bits of it, some parts of the subject, to try to do something more immediately useful. And so for me, the Ph.D. I did that in '76, sometime in the mid '90s, I was doing some things in the pure math side.

But it looked like maybe this could have some value, maybe this could have some value in terms of understanding something, understanding data sets. Now as you probably know, you can, people have ideas and daydream all the time, I do it all the time and never have any follow-through on it or often don't have any follow-through on it. But fortunately for me I managed to sort of talk to my chair into funding one post doc for half a year. And so we got a little bit of a start on making this topology, this abstract subject into something more applicable. And then we grew, we got other funding and so on, developed it into a much larger project, DARPA project ultimately. And it's from that DARPA project that the spin-off came and it's that project where Gurjeet and I work together. And Gurjeet you got your Ph.D. in ICME here at Stanford. Can you tell us a little bit about your path into the Ph.D. program, and how you ultimately met Gunnar? Yes.

I'll try to do it in five minutes or less. So I grew up in India and I was an electrical engineer there. I used to work at this company called Texas Instruments, and that was good. But I had this sense that if you are an, and I think many engineers might relate to this here. If you're an engineer you know math, but you know that you don't really know math very well. You can, maybe you can factorize a metric or something, but that's pretty much the extent of the math that you know. And I had this idea that if I somehow knew more math, then I would be able to do more things in life. And luckily for me I found this program at Stanford. It was called Scientific Computing back then and the department changed its name actually recently. And it was a combination of computer science and mathematics and I hedged my bets.

I said if I don't do well on the mathematics part at least there is this computer science, I'm pretty good at that. And so I applied, and I got to Stanford, with pretty much a quarter's worth of funding that my families are off to great pains to put together somehow. So my sense was it's not a big risk, I land up at Stanford and we will see what happens. If I'm able to survive and find some way of financing my studies then that's okay, if not, life was not that bad in India. So I got here and you tell me if I'm allowed to - I guess I will just say it and then we will see what happens. Okay. Let's go for it. So I got here and I wrote a program to crawl the Stanford website, find out all the professors email addresses and spam them saying I'm this Indian student and I'm here, I can code my way out of many problems. Sign of a great entrepreneur. And it turned out that over the next couple of weeks.

I met with a bunch of professors and there was one in particular in the Aero Astro Department, Anthony Jameson. I showed up to his office and he was working on this Computational Fluid Dynamics problem and I didn't know anything about Computational Fluid Dynamics. I could program my way out of pretty much anything, but I didn't know anything about Computational Fluid Dynamics. So I showed up and I said look I'm this Indian student, I don't have money, I need this sort of work and if you're able to support my education that will be pretty awesome. So he said, he proceeded to ask me like a million questions over the next hour about Computational Fluid Dynamics and of course I knew nothing. So at the end of this interview I'm sitting there with him, he's like "so you don't know much about Computationa1 Fluid Dynamics. I get that you can code and how things are going to work out?" And I had an idea that I was working on at TI back in India, which was around using clusters of DSPs to pull together a super computing capability that you could put under your desk. Because in Computational Fluid Dynamics you pretty much solve very simple problems over and over again. And so, I proposed look I don't know much about

CFD, but I know enough that it's a numerically simple problem that you just need skill to be able to solve. So I pitched that and he said okay so he, I agreed, he agreed to sort of work with me and that's how I basically stayed here for the first year.

We took it to, after we were done building a prototype and some research we basically took it to Boeing Phantom Works. And we met with the CEO and we said look this is really great, you've spent all this time acquiring computational resources which are expensive and this way you will sort of enable all of your engineers to run simulations very quickly. So you should finance some research here. And we got told, you can't really do it, even though we had a bit of a prototype, so I was basically out on the street again. Then I basically started working with Eric Dahl in the mechanical department and it was a similar idea, except it was not Computational Fluid Dynamics anymore, it was Computational Mechanics. And right around the time GPUs were becoming available and the computer power was becoming apparent. So we started, I started using those. So we took it to NSF after about a year and NSF was like "you already have it working, why do you need money?" So I was basically out of financing after another year. Around the same time I saw an e-mail from Andrew Ying in the Computer Science Department and he had this program to build a four-legged autonomous Robo. And I was pretty good at robotics, it was something I had a passion in undergrad, in my undergrad.

So I basically started working with Andrew, we built a prototype of this four-legged Robo. It did really well actually, that worked out pretty nicely. But Andrew didn't really like me, so I guess he threw me out after about a year. Around the same time I saw an email from Gunnar, and he was talking about using algebraic topology to understand large complex data sets. And in his note he was talking about using the same core set of ideas across many many disciplines, in vision, in computer vision, in neuroscience, in cancer research, in protein folding. And that was really exciting to me because it was a new idea of mathematics, I knew nothing about, so I could sort of satisfy my math geekiness there. And in machine learning and working with Andrew what I had realized was that to solve every problem you basically had to reinvent the wheel. If you were going to solve a vision problem, you were going to have to learn about image processing and Gabor filters and wavelet transforms and so on. If you're going to solve a locomotion problem, then you had to go learn about locomotion and inverse kinematics and so on. And here I saw this note from Gunnar, which was about using the same core set of ideas across many fields.

So I wrote to Gunnar, and I was pretty lucky that he agreed to take me as a student. For the next year and a half I basically ate his head off, asking the same questions over and over again. And anyway, that's all of my story into the ICME program. So, Gunnar, you take a three-time failed grad student into your lab, and you basically adopt this orphan grad student. What was different about Gurjeet? Was he different? So first of all let me say that Gurjeet has misrepresented the interaction that we had in that first year and half actually. I don't agree that he knew nothing at all, in fact he was absolutely one of the very sharpest students that I have seen. But you see many sharp, smart students around Stanford and other universities, but there is another aspect of things which is not only you need to be smart but you need to want to do something and to carry something out and to actually solve a problem, as opposed to writing a paper or finding a pretty piece of theory. And that's what I saw with Gurjeet because I started talking about these ideas which, I am a mathematician, they were sort of only formed to a certain extent. And Gurjeet got them right away, and then said "furthermore I want to implement this" and not only that he come back in two days and there would be some kind of prototype going. And so that I found extremely impressive.

And to me it's something that I've now understood. After the fact in pure math that stuff doesn't apply somehow. In pure math, you're just trying to do something clever, if someone is clever, they will do whatever they can do. But now when I look at students I think about a lot of different aspects of what they do. Yes, they can be smart theoretically, but also it needs to be, yes I want to do something, I want to do it quickly, I want to prototype, maybe with the rough and ready prototyping, let's just get out and see how this theory works a little bit in some simple situations. And that's the thing that from my point of view is kind of often missing on the math side. We kind of, and other sciences too, we want to kind of build up the science in a very pretty and systematic way, but without checking at the end whether things are going to actually work and do something. So what I would say is that the better idea might be that, let's do a little bit of theory just make some guesses at it, and then try to see does it actually work or does it get close to working or does it capture anything that we're trying to capture? Well, that's interesting. In electrical engineering and computer science particularly at Stanford you see a lot of start-up companies coming out at the undergrad level, at the graduate level. What do you think it is about math or the sciences that have prevented them from really doing that? Yes, so one of the things you should understand about, and I'll just speak about mathematics itself.

That with the core way that you value contributions in mathematics is often "how hard is it?", or "how technically deep is it?" I think in some of the engineering disciplines that may be a part of it, but the other part is more directly goal oriented in saying "what can it do for me?", even if it's some simple bit of theory or some simple bit of math, if it's useful, let's go do it. And so I think that is a situation, that sometimes we tend to value scientific research the way we would athletics, that is to say how powerful are you, how - what kind of tough calculations can you carry out or what kind of heavy duty theory can you do? I think that is going to change, actually I think you're kind of even seeing it now in some math places. People are getting a bit more opportunistic, and the thing I think about math is that although we have got this way of evaluating things, there is a huge amount of available math out there where one can take rather simple bits of it and try to apply it to a lot of different situations. And people usually don't because the subject is esoteric, so there is only a few number of people who can sort of recognize the opportunities. Well then how did you ... I would just like to rephrase that for all the budding entrepreneurs in the room - there is a bunch of low hanging fruit in math. It's worth you finding some time trying to fish it out. And well, Gunnar then how did you get from this state where in your department it's very unusual for someone to go from this theoretical idea to something that's a little bit more applied and starting a company. How did you traverse that journey? Well, it's interesting. So again my starting point was very theoretical math, even in my teens as an undergraduate.

I did have a father who whenever I would come home for Christmas or holidays, he would ask me what I was doing, and then he said what are you going to do with that? What's the good of that? I'm sure no one here in the room has ever had that experience before, right? But, so he would do that and so it kind of made me sensitive to it. I was also sort of fortunate to have a friend from graduate school, a contemporary of mine who is a co-founder of ours, Harlan Sexton, who after he did his Ph.D. he went off and worked in Navy labs and then ultimately in software start-ups, and finally at Oracle. And so we had a lot of contact and we would talk back and forth, so he kind of knew the math and we'd kind of try to figure out, was there something that we can do. And we did some fun work, interesting, even useful work with certain kinds of communication networks back in San Diego when I was there and so on. And so he was a sort of natural for me to get in touch with, when we were starting to form this company. He had influences in your life? Had influences, yes that's right. Did you ever feel amongst your colleagues or from the other people within either your department or your surrounding area, where you felt like kind of a sellout for heading in that direction, starting a company? Yes, I think so. Actually I have to say not really - no, I was kind of anticipating feeling that. But no they've been pretty appreciative of it.

I think people have enjoyed it and liked it. They like the fact that it's out there. But the other thing I would say is that of course it's not actually being a sell-out. In my view it's kind of like this, there is a whole spectrum where you start, on the one end is the theory and on the other hand is the actual application. And you can choose to work on the theory end here, and maybe move a little bit in the applied direction, and then when you've done something that looks like it could be published in an applied journal, you can declare victory and say and now I've done an application and you go back and do the theory and so forth. The other point of view you could take is, actually I'm not going to insist that the math be the most colossally difficult. But I'm going to insist that there would be something useful. So I start up from this side and I say look, can I do some simple testing things and then gradually grow the sophistication of the math that I apply to it. And I think it's that second part, that realization of that, that I find is pretty powerful. It's kind of like you want to just, you want to be able to prototype things quickly and try things against actual applications.

Can you give us a quick understanding of what topology is? Yes. So topology is the part of mathematics that deals with trying to describe and represent a shape. In fact it's actually a form of pattern recognition. So it was started in the 1700s by a Swiss mathematician named Leonard Euler. And it's actually thrived on the pure math side. What it does is it actually introduces in variants that allow you to measure shape. So measuring shape is sort of a, that's a funny sounding concept. Because shape to me is kind of a, it's an ill formed or a kind of vague notion. And so the idea of measuring it with numbers is a little counter intuitive. But it turns out that there are ways of doing that, actually very, very interesting and powerful ways of doing that.

The second part of it though is, and I don't think it's usually talked about this way, but topology is about compressing shape. It is about finding, if you think of a circle for example, a circle is infinitely many points and infinitely many pairwise distances between those points. Now if you're willing to sacrifice a little bit of detail, like the exact nature - curvature and so on, you can represent it by say a hexagon or an octagon which is say eight nodes and eight edges, which can be represented in a single byte. I mean so it's very very simple. And the notion of trying to sort of combinatorially compass the notion of shape into something much more understandable is the second thing that topology is about. And so in this area, what happens starting in about the year 2000, a lot of people started to have this idea that we should say these techniques in topology, which are about pattern recognition and representing shapes and sort of getting really precise about what it means by shapes, should now be transported from the pure math world where you're dealing with things where you have complete information. You have all the points or, or a description in terms of equations, to something where you only have sampled information, which is really more like real life. And so that's what's been going on in the last 15 years, porting all those techniques from understanding shapes, and I mean here even higher dimensional notions of shape, not just two and three dimensions. Porting them into what we would call the point cloud world which is, that is where data lives. Great.

Now, so then Gurjeet take us through then the history of that. Going from this, four sets of papers that you've written that are more mathematical in nature, and how does that turn into the notion that you're going to build a company? Yes, absolutely. So just to back up, right as Gunnar said around the year 2000 or so actually DARPA and NSF realized that the way people did science had changed. They realized that people had started doing science by creating new large complex datasets. As opposed to the past where people would create confirmatory datasets, people had started creating exploratory datasets. And new science would happen when they discovered something new from those exploratory datasets. And so they had this idea that people who were doing the best science or creating the best datasets were probably not the best people to analyze those data. So they felt that they might draw incomplete or incorrect conclusions. And so they sort of wanted to find a brute force way of approaching this problem. This could you compute your way out of this problem? And so to that end they started financing research efforts in fundamental mathematics, and topological data analysis, the research that we were involved in was one such effort.

So academically as Ann pointed out, we were very successful based on the same core set of ideas. We published in areas as broad as image compression to neuroscience to cancer research, to protein folding. And so by the time we were done with the research we were pretty certain that we could actually have a meaningful impact in the world. We knew that the sort of, the process of taking a more automated approach to taking large sets of data and converting it into insights would have a lot of value. And we knew that we did not want to do it within academia. So we published all of our research, left Stanford, at least I did, with Harlan our other co-founder, and we started building Ayasdi. And when we started out in 2008, we did not have a product. So we had all this research and it was a challenge to see where might we apply this research? So, basically for the next 2.5 years, we did not grow the company significantly. We were still - we remain as small as possible and I met Ann during our first- I guess second-year of operations in 2009, because at one point the three of us basically... We had a decision as to who is going to be the CEO.

And I wanted to be, so yes I was stuck being the CEO. And I realized that I knew nothing about running a company. So I started taking the Stanford ETL lectures. You should pay attention, it actually helps. And we took a class by Steve Blank which was talking about entrepreneurship, that's how I met Ann. And what Steve Blank, the one thing that he said which stuck with us was basically get out of the building. And as soon as we had a prototype ready, we got out of the building. So I wrote another program to spam, Stanford Alumni this time, and we went out and met with anyone in a business setting who would take a meeting with us. So over about a three month period we met with some 40 or 50 odd people in a large number of industries and got thrown out of roughly half the offices. And to their credit actually, those people that we would go meet, we were not the best at explaining what we did.

And so we were going to these meetings and talk about a bunch of math and people wouldn't get anything. But the other half of the meetings were very significant because we came back with a prioritized list of use cases and people would say "oh, if only you could make this work for this problem in my organization, it's going to be worth this much, this is what I spend, these are the resources." And after like three months' worth of searching we had a long list of use cases. Also I remember there was one point where someone said "Wow! I will write you a $50,000 check if you will just give it to me right now." And the point where from a venture capitalist perspective you start frothing at the mouth is, Gurjeet said, "I think this is worth far more than $50,000," and then he left the room. Yes. In fact that contact I actually had met after an ETL lecture. There you go. Your customer may be in the room. Yes. So at that point I went back to Ann, we said "look Ann we've had a lot of fun working with you, I have learned a lot," I certainly learned quite a bit from Ann. And at this point we are ready to actually start pouring some resources into the company and growing it.

I knew what to do at this point. But I mean, to your credit it wasn't just theoretical. There are a few example cases that you had actually created. And so can you tell the audience a little bit more around what were some of the use cases that you showed, both to me and to these people that you were talking to, so that you could show the power of what this math could be? Absolutely. So we had built a few use cases. One of the big ones was in the pharmaceuticals around drug discovery. So we noticed that pretty much every biotech or pharmaceutical company at some point or the other ends up dealing with this type of data called gene-expression data. And the problem that they are struggling with there is it has a small number of samples, very high number of dimensions and you want to discover information from it. You want to discover subgroups, you want to say what defines these subgroups and so on, and our software did that pretty naturally. So that was one of the use cases.

In fact that was a meeting we walked out of, it was a biotech company in the area and they said "look this is great." We would like to buy it, here is $50,000. Then about this is... And Gurjeet is actually being quite modest here because some of the use cases that I saw, there were cases where people would say "this used to take us 18 months, 24 months to actually perform this exact kind of analysis. The knowledge that you gain from our dataset in a day or less than a day was something that used to take us 18 to 24 months." And so that degree of difference in terms of impact that you're having was pretty tremendous. Absolutely. And that's why we walked away from $50,000. I did the math, I said it saves you 24 months of work, there is probably four people working on it and you're paying me $50,000, that doesn't add up. So we had other examples in the financial industry in which you could take sort of index data and you could predict the microstructure of the market as it's evolving. We had other examples in carbon capture of all things, how do you design a molecule that can capture a lot of carbon dioxide computationally. So these are some of the key examples that we would sort of show - show in these meetings.

But so then, Gurjeet, how did you think about when to take venture financing? Because I mean it felt at least from my end you were being very thoughtful about it, right? Yes. Most people that I approach with a $1 million check, they will say sure I will take that. But for you, you were one of the first people that ever said to me well no, I can't take it right now. So what was the impetus, or what were you looking for before you actually went out for venture financing? Absolutely. So for me I have a very scientific approach. So for any action you should be able to say what you're going to do with it. So the first time when we met after the class I had - I did not know how I would accelerate what we were doing at Ayasdi with more money. Like more money did not equal more success at that point in time. And after we had these meetings and we had these chart as use cases, at that point the equation was pretty clear. You use more money, develop the product, the outputs are clear.

So that was of the - for me it was a very structured way of approaching this problem and saying "if we were to take this action at this point in time, then this would result in these outcomes." So that was sort of my- that was my thought process in approaching. And to Ann's credit actually, when we approached Ann for financing, she was like "You don't know much about venture capital and running a company, you should go meet with a bunch of other VCs." So that was all so - that was a very fun time, meeting with a bunch of VCs, getting perspective and just a plug here for Floodgate, they gave awesome deals. Yes, I paid him to say that. Literally. So - literally. Well, so Big Data has gotten a lot of hype, and Ayasdi clearly fits into that sector in some way. So what's wrong with the term Big Data or is there something wrong with the terms Big Data, and how do you describe that market? Yes, so Big Data in my mind is a meaningless term. It has - I will give you an example. I was - I met someone at a party recently, and this guy was building a small business in which they would approach big companies and make product videos. And he was trying to raise financing and so he was picking my brain as to how do you go about it and he said I have an idea, I think I should raise it - like I should spin it as a Big Data company and then I would be able to raise financing.

And so I said how is this a Big Data company? He said "well video files, they're pretty large!" It's starting to - I think that this whole Big Data thing is at a stage where you go to an investor and you just whisper Big Data and they might give you some money. And so it's a meaningless term, right. I think the more, the far reaching impact of this movement, I don't think there is a name for it yet, is pretty large actually. And I draw my inspiration from science fiction. I grew up reading a lot of science fiction and if you read enough science fiction, there are - you see two types of futures that our species seems to have imagined. There is the one type of future in which everyone kills themselves and either the technology runs amok or people are just not that great. And then, that's not very interesting. There is this other type of future in which it is a post-governance world in which you have free time and a lot of work is taken care of with machines and people basically just contemplate the universe. And in that future you don't see people writing SQL queries or you don't see people using business intelligence software. You see a world in which autonomous systems deal with large streams of data and then form your life to make it a better - to make it better.

And I believe that's what we can achieve with the data. I don't believe that Big Data in itself actually is trying to achieve it. Big Data is just database people trying to sell more database software. And where does topology then fit into that? I think, so from my perspective right, topology is something that allows us to marry a very, it allows us to take a very computational approach and it allows us to build autonomous systems with much more ease than previously possible. So, and that said though, there is still a lot of work. What we're doing at Ayasdi is still just the tip of the iceberg and we have had a lot of success doing what we do. What are some of your - what are some use cases that you're most proud of to date? How funny you should ask. We are - I can show you a few. So this is a use case from a bank. And the idea here is that every - this is a network in the computer science sense, in that there are nodes and there are edges.

These networks are created automatically without ad-hoc parameter selection or anything of that sort. Every node in this network is a compression in the sense that it's a group of transactions that are similar across a bunch of characteristics. And two nodes are connected if they actually share some transaction. So you can think of this network as a giant Venn diagram, that's what it is. So in this example, one of the big problems with financial systems is fraud. And detecting fraud you will - you'd be surprised if you go into a large company they will have hundreds of thousands of rules to sort of, that are built manually over a long period of time to flag possible fraud. So in this example we basically took a bunch of transaction data and the idea was, could we automatically detect the failure of the rules engine? And the regions which look - which are colored red in this picture basically show you those regions. Another example, this is from a hospital. It's a triage model. So if you - if someone walks into an ER, usually the practitioner will go over a small set of questions to try and assign you a score, and the score is supposed to be - it's supposed to say are you going to do well or do we need urgent care right now immediately.

So on the top is basically this network in which every node is a group of patients and they are grouped together based on their similarity across all these scores. And what you can see is the color of this network shows the predicted value for what's going to happen. Low is good; blue is good and red is not that great. While if you actually color it by what actually happens, that's the picture below it. It turns out that there are systematic problems in this predictor model, that's trying to predict what's going to happen with people. And inside there is a population that's circled there, which contains people that the predictor originally said are going to be okay, but actually they don't end up doing all that well. Ayasdi in this case is giving a map of that data? That's correct. Right. And it turns out that these people were the people who are too groggy to fill out the form properly. So they just didn't - they were not able to answer some of the questions.

That should have been a feature that goes into building that predictor. Another example, this is also from a hospital. So hospital administrators have this problem that under Obamacare if a patient walks into a hospital, they get treated, but then if they are readmitted in short order, then the hospital doesn't really get compensated in Medicare. Is that Medicare? Well, it could be Medicare but it's also other insurances. Right. And so from an administration perspective, they want to be able to figure out what are the characteristics of doctors. What makes a successful doctor in this hospital system? And so what we find - the region that is highlighted there, are basically two groups of doctors that are very distinct in their patterns. They operate very distinctly for the same disease and both of them give very high prescriptions. Now I can't disclose the details as to what happened there and why because of the confidentiality. But this is one of those insights that is worth $100 million to a hospital.

Great. So after seeing some of these use cases, so it's pretty convincing, product is pretty interesting. But what about the path from, the difference between being an academic which you both were, to then running a company, building a company? Are they similar, are they wildly different, how would you characterize? Well, they're wildly different, I mean in a very interesting way. I found it a great experience to sort of find out what running a company is like. But I would say that the difference is a little bit - is the following. In academics you're allowed to pick and choose the problems that you work with. So you can deal with the problems that you can find clever solutions for and publish them, and that's very valuable because there is information out there about what works and what doesn't. But in terms of the company, no one cares about how hard or how interesting the underlying technology is, it has to solve problems. So you're forced to get much more focused about what you do. And I think that's the biggest difference, but - and that's been very instructive and very interesting for me to find out.

I think one thing that I would add there is, if you at least in the venture capital world and the world of start-ups, academic is almost a perjorative, would you agree? Sure. And that's really weird. You're technology in search of a problem, right? Yes. And we were certainly that. And a lot of companies that have - that start with a bunch of research, are like that. You have - you have a solution and you might not actually know the exact problem, so you might actually have to go out and find the problem. And that's okay, and if you take a scientific approach and you're a rational person and you're not sort of married to your only one idea, that you had in the - early on, then things can work out. It's actually okay to have a solution and then search for a problem. It can be made to work. Great.

Why don't we spend some time taking questions from the audience then? Yes. Yes. Hi. Thank you for the talk. I have - I was wondering if you could elaborate a little bit more in the process of getting people to share their data with you and what are some of the caveats and recommendations that you'd give if we would try do the same thing? So the question is what is the process by which you get people to share their data with you? So, look what we did which really worked very well for us is we basically went out and got as much public data as we could get our hands on. So that basically allowed us to demonstrate the power of our software without needing someone's private data. And we also operate, roughly half of our customers are cloud-based. We operate a cloud-based service and roughly half our customers are on premise. So for the cloud-based customers when we first went into pharmaceutical companies, the average cost of a dataset in pharmaceuticals is in the $0.5 million range. So it's really expensive data and they were not ready to part with it.

And so we said here is some public data, just go play with the software. You already paid, just play with the software and if you like it then we will talk. And within three days people were uploading their own data. So that was a trick that we followed. I would say also on the academic side, an approach there where there is not commercial interest is to offer to collaborate. That is to say people don't like when you say "send me your data and I'm going to do some analysis on it and then I'm going to publish it." But if we say "actually we would like to work with you and see if we can apply techniques, if you can get something new out of it," people are usually pretty open to that. Yes. So you talked about the case of fraud in the banking system, by the customer? Have you ever aimed the software at the bankers to find the fraud that they've got institutionally, that they're not even aware of. So the question is detection of fraud within the banking system, have you pointed it actually at the banks to figure out what fraud they're committing? We have not - some of our government customers are contemplating doing that. But in some sense we don't have any control over that.

We have not looked at that ourselves. Yes. A little bit about in your private work some of the hardware innovations that you worked on, so FPGAs. And then in the talk you talked a little bit about some of the software innovations. Would you say is Ayasdi a mix of software and hardware innovations and in particular, what hardware innovations have you had? So the question is, in terms of Ayasdi, is their innovation coming from the hardware or software or a combination of both? We chose to be a purely software oriented solution. And the reason for that is because we - so when we started out, we didn't know if we'd be completely cloud based, that was our primary motivation. But we were not sure or whether it would be on premise. And if you're going to sell custom hardware for on premise customers, the type of compliance problems that you get into are not worth dealing with. At this point when we have an established footprint and the company is growing nicely. At this point we're actually investing in some hardware renovations.

But even at this point it's mostly using GPUs and using compute offload cards like Phi and so on. We're not building our own FPGA based solvers and I think all that sort. Yes. Just to take a step back from the technology side of things, I was curious when you're dealing with a clientele that's so diverse, from hospitals to banks. As a company when you think about how you want to progress and develop, how do you take into account the needs of such a diverse array of clients and I guess in a sense determine what trajectory by which you want to grow the company, develop the company? So the question is that Ayasdi is actually targeting a variety of different verticals, from financial services to pharma. And so how do you actually take into account the needs of all these different types of customers to really focus the company as well? So I - we register our Series A financing from Khosla Ventures and Vinod Khosla is someone that we work with pretty closely. One of the pieces of advice that he gave me early on, he said that any time you have a technology that's applicable across a lot of verticals, many companies make the mistake of choosing a path very too early. So he said you should go around the roundabout a few times before you figure out which path you want to take. So last year we actually did not care about the vertical at all. We would go into any vertical, we would talk to them, we'd see if our technology was applicable or not.

Our price points are pretty high. So that tends to be a very good filter. This year - going into this year, we have learned a whole lot about where the sales cycles are shorter, where our marketing is clear and so on. So we're spending roughly 80% of our efforts in the verticals that we have learned about and only 20% of our efforts scouting around. So I have a question that the big case Bernie Madoff who defrauded a lot of our investors in this country for almost 30 years. And nobody knows about it until 2008 and suddenly he ran out of money and I follow it very carefully, and SEC sent people to his office numerous times to check the record. But the SEC still couldn't find it - it couldn't find it at all until he admitted it. And now the court comes out and his people said how they fought even the SEC coming to the office and they have practically changed all of the straight number to default this hole in our country for almost $30 billion. So I wonder now with your - with the blessing of your software, could we prevent this thing in the future for everyone? You know this case Bernie Madoff? Yes. So hopefully the SEC will come to you and bless you with lots of revenues.

It makes sense. So it's a very difficult question to answer definitively. Can you repeat the question? Yes. So the question was take a case like Bernie Madoff and whether or not a company like Ayasdi and the software therein can actually prevent something like that from happening again. Yes. So like I said, the answer is that it's very difficult to say definitively. It really depends on the traces that they leave in data. It's certainly possible, I mean certainly in our interactions with the financial community, whenever they've used our software to look at large sets of data, they've actually found interesting things in every case. And looking for anti-money laundering, in looking for fraud, in looking for patterns of organizations in general from the government side. So it is very - it's possible, but I can't say it's a 100% guarantee.

So I'm going to - I'm a Ph.D. in engineering and so I am one of the people who always looked over at the math side and knew that I don't know as much math as they did. And so from my simplistic point of view, when I look at what Ayasdi does and the power of what they do, relative to questions like Bernie Madoff and data analysis that's been done in the past. The power of what Ayasdi is doing and the magic therein is that it removes a lot of the human components of data analysis. And so a lot of times when you see the flaw in what's happened in the past in data analysis, it's because a person got involved and decided that certain parts of the dataset weren't relevant, or they decided to use a certain algorithm on analyzing that dataset. And Ayasdi is essentially removing those two elements. They're allowing all sorts of algorithms to be tested across the board, plus it's allowing more of that data to pass through that analysis. And the power of that is that a lot of times what we have seen in the past, the problems have emerged because people got involved. And the more you are able to use mathematics to then allow us to understand which algorithms do actually work best and get rid of the notion that there is bad data. There is a real magic in what happens on the analysis side.

That's the investor point of view. Yes, back there. Are you planning on releasing any tools, anything that we can use that's free, that's open source or ...? Are you planning on releasing anything that is free or open source that we might be able to use? So there is already a free and open source set of software that's available on the Stanford website. It's not quite as powerful as what we've developed at Ayasdi, but it has all the basics of what we do. If you search for - I'm trying to... Biomapper? Yes. Yes, I can - if you follow-up with me I can tell you where to find it. But it will get you started at least. Back there in the blue. So if you were to receive maybe like hundreds and thousands of documents not necessarily tagged, some have numbers in them, others don't.

And let's say they were all describing a particular industry. What kind of knowledge would Ayasdi be able to get out of this? Like would I be able to answer okay here are where the big opportunities in this industry are, would I have to go through the data and really start editing, like tagging each document ....? Is this text data? Let's say it's text data, maybe some of them have charts, maybe some of ... So the question is you have a bunch of data that might be text, that's sort of industry based, is that correct? Maybe it's industry based, maybe it's in particular topic area. Can you like - is it - can Ayasdi summarize it and answer specifically? Can Ayasdi summarize and analyze this data? Absolutely. So funny you should ask, in one of our early customers was Merck, the pharmaceutical company. And we were in a meeting at Merck at one point in time and there was a lawyer in the meeting who happened to be in the wrong room. But the presentation was visually interesting so he ended up staying for the whole presentation. At the end of the presentation he walked up to me and he said look this is all great for our drug discovery programs, but I have 40 lawyers who work for me and we spend all of our time in Google patterns, trying to search for patterns, like click next and see another 10 patterns, click next and see another 10 patterns. And wouldn't it be great if somehow your technology was able to summarize all of that work and people could easily sort of understand what was going on in all the patterns, all at once. And we came back and we shared the feedback with the team and one of our engineers, I think his girlfriend was away over the weekend or something, so he basically he built a prototype which is what this is.

It's a search tool for pattern spaces. So you search for something, it pulls out all the patterns from the U.S pattern database and it organizes them into this summarized form. Every node here is a group of patterns that are similar to each other textually and I'm not sure if you can see that, they're probably too small to see, but there are labels on the visualization itself, which summarize sort of what those patterns are all about. So a flare here might contain patterns about the display of images on the computer screen. Another region might contain patterns around the interaction of a mouse with a window on the computer screen and so on. So this is exactly what this does. A lot of our government customers use similar capabilities for large corpora of text. One final question. Yes. Can you talk about Ayasdi's work in treating disease specifically for cancer? I've read of some .

Right. So the question is what has Ayasdi done in mapping cancer? That's a great question. It also has a great story behind it. So right, in one of our 60 or 70 odd meetings with people in the industry we had met with Ann. And Ann sort of had sent us around to go meet with other VCs and get some perspective. So one the VCs that we met, Morday Vidal , they basically said this is a great piece of technology, we would like some deeper diligence. So they basically had us meet with the - with this lady, her name is Beck and I met with her in downtown and I was showing her some work that we had done on breast cancer at Stanford. And halfway through, I'm not a biologist, I'm a mathematician, computer science sort of a guy, so halfway through the meeting she stops me and says this is great, but you don't know why this is great. So she basically stopped me and proceeded to lecture me for the next half an hour about the specifics of cancer research and how this was great and we ended up publishing that work. It's out in the open now.

But in general our work in drug discovery, a lot of our work with our commercial customers is around drug discovery for oncology. And the problem there is that it's a very difficult disease and the - obviously, and the way people approach that today is sort of very hypothesis oriented. You would be scared if I told you the number of people who use Excel to research cancer, and that's not a good thing. So, a lot of our customers actually do use our software for drug discovery and the idea is basically understanding subpopulations, the genetic causes of why they are different and constructing diagnostics. Does that answer your question? I think he is under selling it. So what he showed actually in that demo, I actually saw it as well, was that you could actually see spots of patients who had a certain kind of breast cancer where they were told that they would die within a relatively short period of time. And for some unknown reason they actually survived, but that set of breast cancer tumors were actually quite close to people who did pass away relatively quickly. And so that begs this question okay, if they're actually fairly similar, is there a way of turning the more fatal version of that tumor into the one where you can actually survive. It's a - there are some really interesting questions in there that I thought were pretty powerful that Gurjeet is glossing over here, but there are some really interesting questions that you can actually now pose as a result of this type of analysis. And then the last thing I would just say is, and Tina is creeping up here on me, but this is a story of one of our students.

So in Mayfield Fellows Program, we convinced one of the students to actually do a summer internship at Ayasdi and in that period he did this great analysis on basketball athletes and the dataset they're in. And through this summer internship he discovered, what was it, that there is 13 different positions, not just five. He presents this at MIT at some athletics forum and as a result of that he is now in the thirty under thirty list on - in Forbes magazine for ... Yes, two years running. ... two years running and not in high-tech and being analytical or the fact that he is at med school, but rather in sports. Right there next to lots of sports athletes whose names I don't know, but I know they are very famous. So I have been telling all my students who really want to be athletic, well if you really want to be in that list, you don't have to go and do sports, you don't have to be athletic, you just need to be really good at math. On that note, let's thank our incredible guests today.