Actuaries Institute Podcast
Title: Machine learning explained

Date: 15 July 2021
Interviewer: Meg Yang
Guests: Jacky Poon, Michael Storozhev
Duration: 20:46

Meg Yang: **What is 'machine learning', how do we adopt machine learning in our work, whether it's in traditional actuarial fields or an emerging area, and which industries are likely to benefit from it in the future?**

Meg Yang: **I'm Meg Yang, member of the Actuaries Institute's Young Data Analytics Works Group, also known as 'YDAWG'. Joining me today are two of Australia's leading actuaries in data analytics, Jacky Poon and Michael Storozhev, to share with us the ins and outs of machine learning. Welcome Jacky and Michael!**

Jacky Poon: Thank you, Meg.

Michael Storozhev: Yeah, great to meet you. And thanks for having this really important conversation.

Meg Yang: **Thanks. Now Jacky, this is your first time on Institute's podcast. As Convenor of the YDAWG and in your capacity as Head of Finance at nib Travel, do you use machine learning for work and how would you define it?**

Jacky Poon: So, I've used machine learning in my career. I've worked in pricing and analytics previously, and I don't get to use it as much in my current role as Head of Finance, but I'm still really passionate about data analytics and what it can do. I think machine learning is about using automation to fit models more efficiently and more predictively.

Meg Yang: **And Michael, as the Head of Underwriting for Asia and startup ventures at Cover-More. Can you describe your experience with machine learning?**

Michael Storozhev: Yeah, well, there's a lot of hype, a lot of confusion between what machine learning is. It does not help that terms such as machine learning, data science, and deep learning, as well as AI, artificial intelligence, are used interchangeably. In terms of what machine learning is, history tells us a lot on this issue.

Michael Storozhev: So, for me, a pivotal definition is a field of study that gives computers the ability to learn without being explicitly programmed. Surprisingly, this was a definition from back in 1959 by Arthur Samuel, a pioneer in the field of machine learning, who defined the term. But practically for me, it's always

about how much do we need to be involved with a computer as a human. One can always tell a computer how to run a program by giving it all the possible if/else statements we can think of to replicate our intelligence, so however wrong it might be. But machine learning is when the computer starts to create those 'if' statements without our specific input.

Meg Yang: **Thanks, Michael. I love all the history and the quotes. So how did you first learn about machine learning and was there someone or something that really inspired you?**

Michael Storozhev: Before becoming an actuary, I was non-traditional. I worked as an advisor for a federal member of parliament. And as you could expect so much about energy, especially in the election year, it was around winning the election. Working on policy, talking to voters, sending out campaign material, and the success metric was to win. So, after Barack Obama's success in 2008, I remember listening to a lecture in 2009-2010 entitled 'How we use data to win the Presidential Election'. From Dan Siroker, who came over from Google to the Obama campaign to be the Director of Analytics, I was enthralled.

Michael Storozhev: I'd just finished my actuarial studies, and here were the same predictive modeling techniques that I'd been studying, called something different, used in a different way, to win the biggest selection in the world. They were the same things we have studied. And I think my first thoughts were, who is Dan, and is he an actuary? But saying such predictive analytics and machine learning used in this area that I, and many others could see would be revolutionary to ensure that from then on, I kept an open mind about this field of data science and machine learning that was rapidly growing and always causing a lot of excitement.

Meg Yang: **Thanks for sharing with us Michael, that was certainly very inspiring. Jacky, what was your first machine learning project?**

Jacky Poon: So, when I first started, there wasn't so many great resources available for learning machine learning and data science. So, I'm really jealous of the people starting out today. Things have changed so dramatically, there's tutorials for everything, there's so many new packages that are so easy to use, and you can just plug in your data and get the models that you want. My first model was on conversion rates, I was looking at conversion rate dataset and looking at the driving factors of that, and then using that to drill further.

Meg Yang: **Thanks Jacky, I think conversion models are such a good example of machine learning application. So, what are some of the common machine learning techniques do you guys use?**

Jacky Poon: For supervised learning models that I think every actuary should have in the toolkit. First is the well-worn GLM that I'm sure all actuaries are

familiar with as part of their education. Another one would be the decision tree. It's nice and simple, just like the GLM, and I think transparent as well. And it can capture non-linear effects as well. Another one is the GBM. So, in all the data science competitions, the GBM is just so dominant. It performs fantastic, it captures nonlinear effects, and it just works really, really well on tabula data. And finally, there's neural networks, which is the model that underpins a lot of the, what you call AI right now, and it's driving a lot of hype.

Meg Yang: **Thanks Jacky, are there more techniques that you can add, Michael?**

Michael Storozhev: Jacky, good point on mentioning GLMs because it's an instrumental machine learning algorithm yet gets overlooked because of the hype of everything else. As actuaries, we need to do better at showcasing someone our existing knowledge, it might be a GLM, but it's also predictive pricing, the buzzword that we always hear so much about. But to compliment the structured models, Jacky that you mentioned, I'm also really excited about the unstructured models that are starting to emerge. These are the models where a particular outcome, classification or label is not known and instead the machine aims to group similar unlabeled data together using the hidden structures of the features. And it's really important for things like customer segmentation based on purchasing behavior and demographics or even tasks that actuaries are more used to seeing, such as NVIC or postcode clustering.

Michael Storozhev: And I have to mention something that I'm really excited about, and it might be an indictment of the data that many of us have access to, but it's been a really exciting recently to spend a lot of time on data mining techniques, using natural language processing or vision techniques to really try and pull out those nuggets of gold that are in the swamp of unstructured data that's all too frequently put into the too hard basket.

Meg Yang: **So out of so many different techniques, do you guys have a favorite and why?**

Jacky Poon: My favorite is deep learning. I think there's infinite possibilities in where it can go. And the reason for that is because it works a bit like Lego, you can connect the neurons in a number of different ways to predict time series, to predict structured data, to predict computer vision, to predict like audio data or text data as well. I'm really excited about what it can do.

Meg Yang: **Yeah, thanks. I think everyone loves Lego. What about you Michael?**

Michael Storozhev: And to take that point a bit further with a particular example, for me it's all about text, text is everywhere. For businesses in particular, almost every process in a company involves manually reading and understanding of text. Traditionally extracting meaning from language was extremely difficult, not only for machines, but sometimes for humans. Human language is messy, complicated unstructured, sarcastic, a far cry from a highly structured data that machines are so used to dealing with. So,

most of the time text is analyzed by humans who read that text and transform it into structured information. Just look at yourselves, how many emails, reports do you read each day, so because of that, my favorite has to be natural language processing models, NLP. These are pre-trained models built on essentially the entirety of written language.

Michael Storozhev: Imagine throwing all of the internet into a model to do a particular task. But the great thing is that these models have already done the hard yards for us. We don't need to throw all that data into the models each time, we don't need to relearn English from scratch each time. Instead, we can adapt them with just a few lines of code to perform specific tasks such as fraud analysis, or even customer sentiment predictions.

Meg Yang: **Yeah, that's really fascinating. So, Michael do you actually apply natural language processing to your current work?**

Michael Storozhev: Definitely. Not only from data mining that I mentioned recently, such as trying to use free text claims data to extract what happened to a customer, what caused a particular event. In travel insurance we also work a lot with medical reports and doctors' reports. So being able to understand what caused a particular illness or injury is instrumental to be able to get the best possible customer outcome. But most recently, what our team has used these models for is understanding what customers are saying when they go on to websites such as product review, customers there are talking about such a myriad of issues of complaints that it's very hard to analyze each one individually. And so, applying some of these natural language processing techniques allows those comments, that feedback from customers to be very quickly clustered and monitored so that if they're emerging issues, we can very quickly react to them.

Michael Storozhev: When looking at a great example I saw last year was even before COVID hit, that there was more and more misunderstanding of what the product covered and didn't cover. Can I cancel because my grandfather got sick, can I cancel because my cat got sick, customers don't know those reasons and when they're caught out, they will complain. So that was an emerging theme that we'd already seen from customers. And there's a lot of great work by all the insurance companies going on at the moment to address that misunderstanding of what a product covers right now.

Meg Yang: **Yeah, that's definitely very useful and very interesting to explore. So, Jacky, what are some of the common machine learning applications in actuarial work that you have encountered?**

Jacky Poon: I think the first place I'd look would be our bread and butter. So, pricing and reserving, I'm part of the Machine Learning in Reserving Working Party based out of the UK. And we've been doing a lot of interesting research on how machine learning models can be adapted to model out very traditional triangles. And obviously in the pricing space machine learning is already being heavily used.

Meg Yang: **Yeah, thank you very much for that. So, what are the challenges do you guys face when trying to apply machine learning to your work?**

Michael Storozhev: I have to start off by talking about the lack of focus on the explainability and interpretability of models or techniques that we sometimes see. It makes my blood boil when I hear statements such as “the model says so”. To me that's lazy actuarial work, that's lazy data science work. We've all heard about how we as actuaries need to be better to present a non-technical audiences, such as executives, but more recently, we need to be able to be ready to explain much more detail to regulators and even customers what is driving changes. Understanding the assumptions and limitations of these machine learning algorithms is key to achieve this, but luckily that should be the bread and butter for actuaries. It can help set us apart.

Jacky Poon: I couldn't agree more. In the data analytics newsletter, we've been bringing up these news articles of these failures where models have been biased, racist, or doing other nasty things, and there's always good examples to include in every month's newsletter. And I think actuaries can play a great role in informing that discussion. We have the data skills and we have the governance side covered as well.

Meg Yang: **Okay, so how do you think we can tackle those challenges?**

Michael Storozhev: One needs to be the technical side, nothing can beat understanding and being proficient with the tools that you use. And also, being on top of what the emerging developments and research is happening in this field. And as I said before, we need to understand what the limitations and assumptions of using these models before. There's a lot behind say statements like import XGboost or import Bet, just like there's a lot more to reserving triangle, then a mechanical average of development factors. It's essential as that we as actuaries are more comfortable with these assumptions in machine learning algorithms, just like other assumptions in any other actuarial model.

Jacky Poon: I agree there too. I think there's always a temptation to chase the shiny in that neural networks, like, cool, let's do a neural network, but do you need that or could you just have done that with a pivot table. Sometimes we over-engineer things and that can create issues down the line. These things are tools and we should use the right one for the job.

Meg Yang: **I can't agree more. I think actuaries are really well positioned given that in our education, we've learned how to contextualize, how to solve a business problem, and what to focus on when making a decision. Machine learnings are essentially just tools that we can apply. But eventually it's us making that judgment call in terms of what's the best solution for a business problem. So just out of interest, which area or industry do you see machine learning get applied more and more often in the future?**

Jacky Poon: I have said all of them. We're just living in an increasingly data rich environment, and now even your fridge is full of sensors in them that connect to the internet and you'll be able to do analysis on that data perhaps. So, I think data analytics will touch on every area and if they aren't already. And what we do with that is up to us.

Michael Storozhev: And the danger of having so much data, the danger of having a supercomputer in our pocket at all times is that sometimes there's too much data. I was reading a survey from Kearney recently that talked about 65% of executives couldn't explain how their AI models make decisions. And these were chief analytics, chief AI, chief data officers. These are the executives that are responsible for deploying and using these models. And more worrying for me was that the same executives said that 73% of them struggle to ensure that there was enough ethical or responsible AI practices, and I'm worried about that. In this world of move fast and break things that we so often hear about the end in itself cannot justify the means, the right controls are essential.

Michael Storozhev: And just look at the negative publicity that Lemonade and InsurTech in the US received just a few weeks ago from the innovation in machine learning that led to, and I quote from Twitter “serious accusations of bias, discrimination and general creepiness”. I want to say that public trust innovation built on machine learning to support all these incredible opportunities as Jacky said, exist in all industries. But that can't happen without more rigor, greater awareness of the negatives possible in the models and great about this. And it's why I want to see more actuaries talking about exactly these sorts of problems, getting involved in these sorts of discussions so that we can be part of the opportunity to drive solutions, to reduce that 65% number, that I'm quite horrified about.

Meg Yang: **Thanks, Michael. I think that's super interesting. We're definitely well-placed to help leaders to better understand the machine learning techniques that can be applied and also to help them leverage the data that we currently collect. Jacky was there something else that you were trying to add?**

Jacky Poon: I guess in the movies, AI already has a bad reputation. You've got Skynet, you rarely see AI doing anything good. And so, I think the industry needs to build up trust in analytics and in AI and in machine learning and what they can do. And I think having the right controls in place is quite key to that.

Meg Yang: **Yeah, indeed. Well, thanks for sharing. Just last question, because I know some of our audience could be someone who has never really done much modeling or doesn't understand machine learning techniques very well. So, do you have any advice for people who want to start their first pet project using machine learning techniques?**

Michael Storozhev: I have to echo Jacky's point about being jealous of all the amazing resources that those that are interested in this field have available to themselves right now, both from the Actuaries Institute and more broadly, but I would always start with something that you're interested in. And I always love giving this example, but a colleague of mine when he wanted to get his hands dirty with machine learning, try to see if he could outsmart the betting odds, break Tabcorp or someone like that. And he did that by targeting a very niche mispriced segment in NRL games, the 40/20 odds, and build a great machine learning algorithm, looking at all the players involved, looking at the time of the day, who, what, why it was happening, and in assessing could really show that he could make a killing off this idea.

Michael Storozhev: Now, unfortunately for him, I think the betting agencies caught on that they were doing something silly because they cut off exotic betting odds just before the season started, so he couldn't implement this amazing model, but it got him excited and involved in this space, doing something that he was quite interested in passionate about.

Jacky Poon: And in the same vein, I'd say practical experience is the best way of learning to do something. There are these fantastic tutorials out there, but nothing beats practical experience. So, I go out there, solve a business problem, something that you're passionate about or something at work. Start simple, there's no need to try to move mountains when you only have a shovel. So just try to do something that perhaps in an area that you're already working on and go from there.

Michael Storozhev: And feel free to reach out to either of us. Because I think I really enjoy seeing more and more actuaries involved in this space. So more than happy to share some of our own learnings and failures in our journey.

Meg Yang: **Thank you so much, Jacky and Michael. Unfortunately, that's all we have time for today. It was a pleasure having you both on the podcast. Listeners, we hope you enjoy this discussion. You can listen to more Actuaries Institute podcasts by visiting the podcast section of Actuaries Digital or via all mainstream podcast apps. If you'd like to continue the conversation with us on social media, please follow us on Facebook and LinkedIn at 'Actuaries Institute' (@actuariesinstitute) and our Instagram and Twitter at 'ActruariesInst' (@actuariesinst). Please write in with your questions and comments on the show, we'd love to hear from you. Bye for now.**