TECH REIMAGINED
AN ENDAVA PODCAST

Predictive Analytics Reimagined - Part 1

Tech reimagined. Redefining the relationship between people and technology, brought to you by Endava. This is tech reimagined.

[00:00:11]
BRADLEY HOWARD (BH): I'm Bradley Howard, and today's episode of tech reimagined is focused on data, what we can do with it today and how it's likely to change in the future.

To help me explore the topic, I'm joined by two special guests, Keith McCormick, Machine Learning Consultant and LinkedIn Learning Author, and Dan Pelos, Lead Data Consultant at Endava.

Keith, would you like to introduce yourself.

[00:00:34]
KEITH McCORMICK (KM): Thanks, first, for inviting me, and I've been building predictive models of one form or another for about 25 years, and the first half of that was building the models themselves, and now I help companies build their teams to do the same thing.

[00:00:49]
BH: Thank you and welcome to the show, Keith. And Dan.

[00:00:51]
DAN PELOS (DP): Thanks, and hi, Bradley. Yes, I've been Lead Data Consultant for Endava, I've been at Endava for around 18 months now, but I've spent the last 20 years working on all kinds of data and analytics projects from a number of different companies spanning across pretty much every industry there is. I've worked directly for software vendors, building out data products but also for marketing services vendors building data solutions.

[00:01:15]
BH: Thanks, Dan. It should be added that Dan and I actually sit opposite each other in the office most of the time in normal times. So how do both of your roles differ, and how would you work together on a project? Let's start with you, Keith.

[00:01:29]
KM: Well, I think the trick to understand where I'm approaching these problems is that I focus very much on the predictive aspect of the problem. So in almost all cases I would imagine that I would be drawing upon the same data that Dan might be, but by focusing on making some kind of a prediction, whether or not a loan is going to be defaulted, whether or not a particular piece of equipment is going to fail, or various things like that.

[00:01:55]
BH: And, Dan, what' your view on that?

[00:01:56]
DP: Whereas with myself, I'm much more of the business versus technical focus. So I work with customers to understand whereabouts they are with their data, the business requirements, and how data can be used to support and grow their business forwards.

[00:02:11]
BH: OK, well thanks for setting the scene, both of you.

So words like data, business intelligence, data science, predictive analytics, insights, micro-decisions, they're thrown around a lot, so where do your roles fit in with these things, and how would you describe the value that you bring to an organisation? Over to you again, Keith.

[00:02:30]
KM: Well, you know it's tricky because this is all on shifting sands. These terms are forever changing. But, first of all, I guess data science, the best way to describe that is it's a very broad term, and there's a lot of struggle really trying to figure out what data science is. But with business intelligence and predictive analytics we can get a little bit more specific. The way that I usually think about it is that business intelligence is very much about the present tense, what is currently going on in the business that allows us to run the business day to day. And predictive analytics, we're focused on the future at some point and I think that's going to be a real theme of the, you know, our whole talk here, probably, today, is what kinds of things are we predicting?

But that's really what separates that. So if you think about the analytics maturity curve that everybody loves to talk about, sometimes it seems to devalue the present tense but, of course, you can't possibly do that because you can't run your business without knowing about what's going on. But with predictive analytics you're very strategic about that handful of things that you really need to predict at a micro level. And I'm sure we'll talk a lot more about that.

[00:03:38]
DP: And from my side it's much more preparing that data to enable them sort of data science activities to take place. So I would work with the customers with their businesses to understand their data and see if it's fit for purpose, see if the information is there that can drive and deliver all of these data science capabilities.

And alongside that I'm working much more with the understanding that data journey as we prepare for enhanced data science or visualising practices.
I also look to identify quick wins along the way because I love problem solving so I like to try and think outside the box and spot other opportunities related to data as well as those initial requests.

[00:04:20]
BH: And when you talk about those problem statements, are those very definite business problems and that you're looking for quite concrete results?

[00:04:30]
DP: So often they have that in mind and they go into a data project knowing exactly what it is they're looking to achieve on the outcome. But quite often the data's not necessarily there and prepared and ready to enable that sort of activity to take place. So it's also looking at the data, looking under the hood, to see if the quality is there to enable these sort of activities to take place.

[00:04:53]
BH: And, Keith, are you always looking for a definite outcome as well?

[00:04:56]
KM: Yes, but defined in a different way. So, you know, Dan raises an interesting issue. It's definitely been the trend in BI for some time now to empower the analysts and the different lines of business to be able to do that on their own. Whereas in predictive there's been a real debate about do specialists do it or should it be done within the lines of business? And that's still being sorted out.

So when I'm trying to nail them down on an outcome, it's usually highly specific to the kind of prediction that's being made, and, ideally, there's only two outcomes, like something's going to happen or something's not going to happen. And I know that when I'm working with executives, if we can get that specific then I know that I'm on the right track.

[00:05:38]
BH: So when people think of predictive analytics they probably imagine that they will understand with a degree of certainty what's going to happen in the future. How do you use and approach predictive analytics, Keith?

[00:05:51]
KM: Well you know it's interesting. That's certainly true. But most executives will think about those predictions and that range that you described in the terms of forecast. They're going to be thinking like forecasts, like what will our sales be next month? And usually what traditional machine learning is much more effective at doing is making a prediction about an individual customer.

So you don't get those ranges like you do with something like a political poll, that's not normally the kind of uncertainty that I'm communicating about the mode, it's usually what's called a propensity score, which sounds technical but it's really a simple thing. It's something like the likelihood of default on the Smith family loan is .87.

[00:06:36]
BH: So in your world you really are looking for absolute outcomes to provide the business with.

[00:06:41]
KM: Well, you would think and there's certainly some truth to that, but what ends up happening in practice is that people initially think, oh, well that propensity score is above .5 so we're going to take some action. But the thing is there are constraints when you run a business, there's only so many people that can get on the phone and call the Smith family about refinancing their loan or something like that.

So what you end up doing is taking those propensity scores and combining them with business rules, and it's that combination that allows you to leverage the predictions. So you might only be intervening on loans that are above .85, and then you might find out that you have a little bit of extra manpower to put on the phone, so you lower that to .8. So it's not just the prediction it's how the predictions interact with other business rules within the organisation.

[00:07:34]
BH: And do you work with the business to actually define some of those operating models?

[00:07:39]
KM: Absolutely. Yes, that's it, absolutely key part. You know a lot of times I think particularly young data scientists think that when you build the model and you're generating these scores that the job is done, and, of course, you're about 80% done at that point but there's still a lot of work to be completed, and that includes to how to get these things to work together. And, you know, one of the stereotypical problems sometimes is that sometimes folks in sales will have models that are predicting things but they might be drawing upon their own personal knowledge, and might tend to ignore those predictive scores.

If that's happening, of course, the predictive scores can have value, so there's actually a lot of cultural stuff going on where you have to sort out how is it actually going to be used in practice, and, is it recognised as useful by the folks that are supposed to benefit from it, because if they don't embrace it it's not going to do any good.

So a gifted modeller, in my opinion, will be working with those folks throughout so that you understand their needs, and only through understanding their needs will they develop trust in the model, which will cause them to use it.

[00:08:52]
DP: And from my side it's very, very different. We are allowing business users access to the data so they are able to view and slice and dice the data in order to understand their own data themselves, in order to spot opportunities for the data and to create insights, and from there they are able to make business driven decisions at a business level rather than really needing to know the nitty gritty and highly technical skills that are required for data scientists.

[00:09:24]
BH: Which brings me on to something else, which I've been wondering about just before we went to start recording this. So when I started working in IT we had a DBA or a database administrator, and now we have whole data departments. Why do you think there's been such a revolution into the whole department philosophy around data?

[00:09:44]
DP: So to start, obviously data's grown massively over the recent years, so data is now becoming a massive business asset. Data is enabling businesses to grow, to make more revenue, and to provide better services to their end clients, customers. Not only that but also data is now a dependency within a company if we start looking at the governance where they have to be able to monitor and track data points throughout the whole organisation.

Alongside that, data is also now rich enough to be able to trust. Previously, organisations have been using their brains rather than letting the data tell the facts. So an organisation might say, I think that if we put this sale on then we will increase the number of sales by X percentage. But now they have enough data to be able to look at patterns, so they can now say I know that if we put on a sale we will increase number of sales because the data has shown us this year on year. So the patterns are outweighing … the data is outweighing the brain.

[00:10:47]
BH: Do either of you have any recent learnings that you've learnt from a recent project, or any anecdotes about predictive analytics or data points?

[00:10:57]
KM: You know, Dan's comments have kind of inspired me to think of a particular anecdote. I was working with a manufacturer and they were trying to predict delivery time for certain parts, and this was really a big issue because if you've got some earth mover or crane or something like that and it's down, it's not just that the equipment is unavailable, that whole construction project has now come to a halt.

So this could be thousands or tens of thousands of dollars a day. But, as we started to look at the problem, what was really going on is the folks that allocated the equipment didn't have very good access to stock on hand. So it became pretty clear to me, quite quickly, that they really weren't trying to predict anything, it was that the client facing team didn't have insight into stock on hand. And when it's not available there's no way that it's going to get shipped on time, and, therefore, it's not going to arrive on time; if all that makes sense.

So where this anecdote takes me, which I think is terribly important, is there's been so much excitement about predictive that folks aren't concentrating enough on the terribly important BI work that Dan is describing, because that's not a situation where they really have to predict anything, they have poor information flow, and because of that poor information flow they're making bad decisions.

If they simply knew what was going on in the present they would have been fine. They don't need to know what will happen a month from now in that particular instance.

[00:12:31]
BH: So in that example, Keith, and thanks for sharing it with us, when you were called in to help understand about the delivery schedules then how do you then broaden that approach to say, guys, can you share some of your stock inventory with me and some of the sales approach, how do you broaden it so much?

[00:12:49]
KM: You know it's funny, when I work with clients and we're in a meeting, you know, room, and typically it starts with just the data scientist, I'll often suggest that we just get the speaker phone, now, of course, I'm remembering, you know, not 2020 so much, for the last several months we've all been remote, but, you know, when we used to be in conference rooms. I'd often say why don't we just call the warehouse, you know, and get them on the speaker, you know.

And I'll often go in that direction. So I try to be very hands on in that way, and that's usually how I'll discover it is I'll want to talk to the person who's requested this. Now Dan pointed out that you also can empower those individuals and have them figure this out on their own, but, as you can sense, the kind of projects that I'm describing sometimes get complicated enough that those folks start out on their own and then they say, wow, on this one we want some help.

So if you simply ask that person, how are you handling the problem now? The secret often lies in their answer.

[00:13:49]
BH: Right. That's very insightful in itself. And, Dan, do you have any anecdotes?

[00:13:54]
DP: Only that from my side is that businesses want to be able to shift employees away from what they were doing, what they've been doing over the last 10, 20 years. A common word that comes

into mind is Excel, so people want to move away from allowing end users accessing Excel spreadsheets that they've been monitoring and changing. They want to be able to put it into better tooling to give better outputs to connect into the same dataset. So everyone's all using the same dataset that's more precise, it's richer, and enabling them to use, report and visualisations and analytics through better toolings than kind of what people have been using over the past 10 years or so.

[00:14:41]
BH: And, Dan, do you think that the modern suite of tools around business intelligence have helped organisations, the fact it's been democratised, the fact that it's really easy to build reports on products on Power BI for one example, do you see that as positive, or do you think that there should be data departments that are still generating that centrally?

[00:15:05]
DP: There's certainly a positive impact in enabling end users that don't need to know the nitty gritty of the underlying, the data, creating calculations, creating reports in Excel, which is very row and column based. But these tools, such as Power BI, enables end users to create very nice, quick visualisations, and being able to give quick, results quicker based on the data because of the technology that it's embedded with within such tools.

[00:15:35]
BH: And to both of you, what impact do you think that COVID will have on predictive analytics? Do you want to start, Keith?

[00:15:42]
KM: Well it's hard to begin to, you know, list them all, but, you know, we're all feeling the effects of work from home, mostly positive in my case, I have to admit, you know, I think I've been quite productive this year. But something that folks aren't talking about yet that will probably be a more long-term impact is how work from home will ultimately impact hiring.

So when I work with a bank or an insurance agent it's usually in big cities, and they haven't had, everybody talks about data science shortages but, truth be told, there's not a data science shortage in New York, I don't think, you know like they can always find good people.

But when I work with regional companies, maybe more like the manufacturer that I just mentioned, and so on, sometimes they're in areas that undergraduates who just graduated from a, you know, nice college in a big city, they don't necessarily want to move to a rural area to take that job with the manufacturer. And it impacts their ability to attract the best candidates. They ultimately find them but their searches are twice as long and, you know, it's a struggle for them.

I think now if we get accustomed to work from home, people are going to consider candidates that they couldn't have considered just a year ago.

[00:16:56]
BH: Right. And do you think that the work from home environment, is that harder to do in both your positions with data because of the confidentiality around the data in itself, even when you're both building models and having access to such a broad amount of data, is that high risk for companies?

[00:17:13]
KM: Yes, I think that's going to be an issue. I haven't had to kick off a new client relationship in the last six months because I've had on-going client relationships, and those on-going relationships have not been affected negatively from work from home. However, you know, one of the anecdotes that I shared earlier about just a spontaneous decision within the context of a meeting to bring in another person, that's easier to do when that person is down the hallway.

And you're absolutely right that data access is part of that as well. I'm not quite sure where that's going to take us but I think that is going to be a challenge, because I have had some clients over the years where I've had to do all the work on site because I wasn't able to access it off-site. And some clients just simply have data that is sufficiently proprietary or necessary to protect that you get into issues like that. And I don't think anybody has sorted that out yet. I think that's going to be a challenge.

[00:18:15]
DP: I think the whole cloud hosted environment might help with this, Keith, I'm not sure about your thoughts, but allowing people accessing data via the cloud through all of the cloud security might help this.

[00:18:29]
KM: I would tend to agree. I think that when I haven't been able to access things off-site, and of course I've been doing this for a number of, you know, years, a number of those years predating, you know, the cloud. But I think you're right, Dan, that when organisations have the team dealing with this because the team is working from home, then the challenges that I've experienced as an external resource will be diminished because they will have already have taken care of that.

Whereas if they've got, you know, a half dozen data scientists and one external resource in the form of me shows up, they haven't wanted to do any kind of exception process. But you're right, I believe that we're going to move towards it not being the exception, it's the way everybody does it. There's so many things like that happening this year.

[00:19:16]
BH: Dan, do you think that COVID has improved the perception of statistics and data?

[00:19:20]
DP: It's a good question. I think, really, I think more people are aware that data, and the science behind it, is being used more so to make decisions on people's lives at the moment, so obviously whenever the news is on we're seeing graphs and charts of what's been happening, historical events. So I think the perception is very much that the data is being used to make these business, these driven decisions.

[00:19:49]
KM: I think, you know there's a couple of things particular to COVID that I think maybe folks are just somewhat through osmosis, it's just kind of out there, that they're absorbing a little bit that I think is helpful. One is the notion of a lag, which often seems like a technical thing but it's really not. I think people are just starting to think in terms of the fact that if there's a spike in cases that you can anticipate that two to three weeks later that there's going to be a spike in hospitalisations.

And that might seem like a minor thing but that plays such a major role on so many predictive analytics projects. So I think folks are developing an intuitive sense for things like that that perhaps they didn't have before.

And I think the second one is this notion of kind of like a what if scenario. We've struggled somewhat in the US, you know, really reaching that point where new cases was low. So there's a lot of scenarios where people will say we think the number of cases in the fall is going to be a certain number if these conditions are met, and this other number if these other conditions are met. These, again, are everyday things to anybody that builds predictive models but it's becoming part of our language now.

[00:21:05]
BH: We're now going to move on to a quick fire round. I'm going to ask you a simple question and I'm going to ask you for a very short and concise answer, if possible, please. We'll start with Keith on each of the questions. So, Keith, what did you want to be when you grew up?

[00:21:23]
KM: Well, when I was in middle school I was really fascinated with being an architect, and I still find myself collecting architecture books.

[00:21:31]
BH: Do you mean a building architect?

[00:21:33]
KM: Yes, exactly, kind of like a Frank Lloyd Wright type, I'm a big fan of his.

[00:21:36]
BH: Oh, great. I wanted to be an architect as well when I was younger but no one ever asked me these questions. Dan, what did you want to be when you grew up? I think I know what the answer's going to be.

[00:21:47]
DP: So not quite an architect. So I grew up playing guitar in the Brit Pop era, so I really wanted to be the next Indie rock 'n' roller, and, yes, I had all of them kind of haircuts, and I was quite often picked out sitting in a busy working office saying, what's that rock star sat there?

[00:22:05]
BH: Well, I think we're going to hear more about that in the second part of this podcast. So, Keith, complete the sentence: Data helps business to …

[00:22:13]
KM: achieve the ROI.

[00:22:16]
BH: And Dan.

[00:22:17]
DP: I would echo that, but yes, to better understand the organisations and their customers.

[00:22:21]
BRADLEY HOWARD (BH): Keith, what's the most interesting statistic or nugget of information that you've discovered on a project, that you're allowed to share with us?

[00:22:28]
KM: Mine is an interesting one from a fraud project. I discovered when the calls came in that they were using caller ID, sounds like, you know, a minor thing, but a lot of times the number that they put on the form was not the same as the number that they used to call and complain and say where's my money. And knowing both of those phone numbers turned out to be a key to the project.

[00:22:52]
BH: Wow. I'd like to hear more about that another time. And Dan.

[00:22:57]
DP: OK, mine is nowhere near as cool as that. I think I would say, well, a lot of pizza gets sold on Tuesdays. So a company came to us and said, we're selling lots of pizza on a Tuesday and we don't know why. And our task was to answer that problem or question. And we did but it was really that the way we did it was fantastic, the way we generated and brought down lots of different data sources, lots of free data sources out there, in order to be able to answer that question correctly, was really fascinating at the time.

[00:23:30]
BH: Can you share with us what the answer was? Was it to do with a television show or sports event or something?

[00:23:36]
DP: It was single parents that were taking children to activity clubs, football, netball, or whatever, so they would sit in their cars, and ordering from a mobile device, that was another data source that we were able to identify, people were likely to be outside their house. So they were ordering pizzas while they were sat in their cars because it was quick and easy at the end of a football session or something.

[00:23:58]
BH: Oh, that's really interesting. Keith, what's your favourite phone app and which app do you use the most?

[00:24:04]
KM: This probably won't be a very interesting answer, but I am on Audible, which is the audio book app, at least an hour, hour and a half a day. So I'm someone addicted to audio books.

[00:24:17]
BH: And Dan.

[00:24:19]
DP: I quite like quick nuggets of information so I'd have to say Flipboard, which is a newspaper article where it collects lots of different articles from various different websites, and collates them into one, and you can create your own mini magazines and topics you're interested in. So it just constantly gives you nuggets of information and if you want to read more you click that article and it opens up, it finds the relevant article on the web and you can read it.

[00:24:45]
BH: Yes, it's a really clever app, I've used it before. Keith, if you could invite one person to dinner, dead or alive, who would it be and why?

[00:24:53]
KM: Well, I was a philosophy minor, if I was independently wealthy I would, I'd probably be pursuing a PhD in philosophy somewhere. So I'd be tempted to say Wittgenstein, who's a kind of obscure philosopher, but he was apparently kind of a moody, unfriendly person, so I think my answer is Alan Turing, I'd be very curious to get his take on the Turing Test this many decades later.

[00:25:16]
BH: Wow, that would be fantastic. And Dan.

[00:25:19]
DP: I'd probably have to go for a musician, so it would probably be John Lennon. The Beatles were my first love and I think John Lennon … I think we'd have some good conversations on music, the way of the world and how it's moved on since his passing, and I think he'd be quite interested in some of the data knowledge, I'm sure.

[00:25:39]
BH: You know we've asked this question probably 20 times before, so to date, and I don't think we've ever had a musician, so John Lennon is fantastic as the first musician.

Well, thank you to Keith and Dan for a great discussion today. In part 2 we'll be finding out how Keith and Dan both started their career path in data, and how Dan manages to balance work and being a rock star.

If you enjoyed this episode please give us a like and subscribe for more.