MODULE 1 UNIT 1
Lesson Video 2 Transcript

IN COLLABORATION WITH getsmarter™

Module 1 Unit 1 Lesson Video 2 Transcript

Practical application of machine learning within a business context

KOSTAS KALOGEROPOULOS: I'm very pleased to welcome Dr James Abdey. James is an assistant professorial lecturer in the Department of Statistics, where he teaches undergraduate courses in mathematical statistics and quantitative methods across the LSE, as well as more advanced courses in market research. His main research interests include market research techniques, forensic statistics, and the interplay of statistics and the law. In this video, James join us to share his experience regarding some of the practical applications of machine learning and how they solve specific business problems.

Where do you believe machine learning is most useful in business?

JAMES ABDEY: Businesses are all about making decisions and decisions have to be made in the present. And these will inevitably involve uncertainties. So, really, the applications of machine learning, if we really simplify it, we can think about attaching probabilities to particular outcomes. Great example would be e-commerce. So Amazon, Alibaba in particular – e-commerce giants. And they'll be monitoring, of course, what people are buying, but not just what's going in their online shopping basket, but what other pages they are looking, and how quickly they're putting things into their shopping basket.

Now, of course, Alibaba are very happy with those sales revenues, but it's not just the stock they're shifting and the sales proceeds, it's effectively this trove of data that they've collected – not just what people bought, but who were they? What demographic information we know about them? What other products they were looking at? And effectively using machine learning techniques to reduce things to simple probabilities.

So what's the probability that a specific individual, based on their demographic profile, may be interested in buying products X, Y, Z. And, hence, if you make those – hopefully good – probabilistic estimates, you would sort of push and promote those particular products to those individuals. Never anticipating 100% success, but, of course, we are trying to maximise the probability of success and that's where machine learning can really help businesses.

Why do you think traditional learning statistics might be insufficient in solving the challenges of modern society?

ABDEY: Classical statistics, if you think more traditional statistics, arguably invented in the pre-computer era, looking at analysing small data sets. Well, now we have vast data sets. So, given the vast data sets we have today, we need advanced computational techniques such that we can effectively dimension reduce. So, think of it like a huge Excel spreadsheet – vast number of columns, vast number of rows – and it's likely, in practice, many of those cells in the spreadsheet are empty. And you could have this problem of large data sets, yes, but quite sparse data sets.

So the challenges in the modern era are how we can have fairly computationally efficient methods to reduce these vast data arrays into something more manageable. So that's not

to say that classical statistics have no value today. I think it's they're complimenting each other. If we consider your elections in various countries, your opinion polling is still done conventionally. You take a random sample "ish" of about a thousand voters. Who are you going to vote for? Which party? Which candidate? And then come up with sort of the classical statistic estimates of voting intention, and the popularity of parties and candidates. So it's not to say that the era of the classical statistics is over. Of course, we need to be mindful of those vast data sets out there, and it's really trying to blend the two.

How are statistics and machine learning used within a legal context?

ABDEY: For myself, I was doing a consultancy project with the UK Home Office. They were preparing a case for trial and wanted to try and convince the jury about the probability of a particular set of circumstances happening in practice. So, I mean, the classic decision-making is that of a jury. And they are making a binary decision whether the defendant is guilty or not guilty. The prosecution will make the case about why this set of circumstances is so improbable, trying to convey to the jury, this person is guilty, and then the defense will argue why, perhaps it was just circumstantial evidence, coincidence, and why their client is innocent.

So I think there's huge scope to support the decision-making of juries, not saying exclusively on a data-driven justice, but I think there's scope to support the qualitative evidence at trial with these sorts of statistical evidence.

How is machine learning changing the way that businesses gain insights into their market research?

ABDEY: Market research, it's all about researching the market, understanding the market. So basically understanding your customers. Now, are you a B2B or B2C company? In any case, you have a customer. It's just, is it an individual like you or I? Is it another business? But if we take, perhaps, the classic B2C at case: what data do we want on our customers? Suppose we need to distinguish between the behavioural and the attitudinal.

Now, behavioural data can easily be done by sort of direct electronic observation – the Alibaba example. There is a digital record about which products this individual bought. But the behavioural data does not necessarily tell us why someone bought a particular product. So I think there's still scope for more traditional market research methods, such as focus groups. You need to— it should not be a purely data-driven, quantitative world. We need to combine the quantitative insights with the qualitative insights, and it's only really by having a focus group to sit down with a small group of homogenous individuals and really talk to them and understand why they have bought a particular product.

'Course, it's not restricted to the commercial arena, the political arena too. I think good political parties are going to be very data-driven. They're going to, you know, their digital, you know, political campaigns will no doubt be hyper-targeted, not, of course, trying to persuade someone to buy a particular product, but, in that case, to put the cross next to the right candidate on the ballot paper. But I think it's a blend of the, you know, trying to predict which kinds of policies may appeal more to certain demographic groups than others. So the quantitative insights, but still, I think, great scope for the use of focus groups in the political arena too. So those traditional market research techniques to really understand the why to gain those qualitative insights, I think they will still be there and, again, to compliment with the quantitative analysis.

Could you elaborate on the more unconventional applications of statistical machine learning?

ABDEY: There was one somewhat flippant, trivial one, which I saw a while ago looking at trying to ascertain the true authorship of some of the Beatles tracks. Okay. So the various Beatles tracks: who penned the lyrics? Okay. And, you know, Lennon/McCartney kind of thing and, okay, Lennon/McCartney would argue, I contributed this amount versus this amount. But then using machine learning techniques to try and see, is there a sort of, you know, a pattern you know of the Lennon lyrics versus the McCartney lyrics. And when the authorship is disputed to try and attach probabilistic estimates to see does this phrase, lyric, whatever, does it seem more probable it came from Lennon or McCartney?

In your copyright issues, let's say some singer has produced a song and some people think, actually that sounds a bit similar to another track by someone from a few months or years ago. At what point can you deduce, you know, where was the originality and should there be some royalty payment to the original recording artist? So that's a interesting case.

If we take the art world, to determine the true provenance of a painting. You know, was this genuinely painted by Picasso? Or was it some very good copy? So is it, perhaps you're looking at the brush strokes? I mean, I'm no art expert. I know what I like, but, if I look at a painting, I'll think, yeah, that sort of has, I don't know, the Picasso style to it. But with the human eye, I think there's a limit as to what level of detail we can go into. Potentially, and I have no idea to what extent this is done, but to try and ascertain from even like the angles of the brush strokes on the canvas of the painting, which potentially might be a Picasso, and compare that with a painting where the provenance is truly known – this was definitely done by Picasso – and see to what extent there is similarity.

What do you think are the next breakthroughs in the application of machine learning?

ABDEY: I would hope, at the very least, in the coming months, years, decades – hard to put an exact timeframe on it. I would very much hope that governments look more at evidence-based policymaking, data-driven policymaking. There are discussions, here in the UK, you know, post-Brexit – the withdrawal from the European Union – that the UK should become more of a hub for data science and, in particular, for the public sector to base policy around data-driven decision-making. Of course, there are ethical considerations: you know, the facial recognition, you know, when is the state collecting data on you, when does it becomes sort of data snooping and there are the privacy concerns.

So it's a big mixture of stuff. I mean, yes, you know, the academics are focusing, and the researchers are focusing, on coming up with sort of better machine learning methods, but one should not neglect the legal, the ethical, considerations, alongside. Will there be a sort of public backlash against all of this, you know, data collection? Who knows? But I think those are possible risk factors in the months and years ahead.