Data science is a growing field. Here's how to train people to do it

February 12 2019, by Riaan De Jongh

Credit: AI-generated image (disclaimer)

The world is inundated with data. There's a virtual tsunami of data moving around the globe, renewing itself daily. Take just the global financial markets. They generate vast amounts of data – share prices, commodity prices, indices, option and futures prices, to name just a few.

But data is of no use if there aren't people able to collect, collate, analyse and apply it to the benefit of society. All that data generated by global financial markets gets used for asset and wealth management – and it must be properly analysed and understood to inform good decision making. That's where data science comes in.

Data science's primary aim is to extract insight from data in various forms, both structured and unstructured. It's a multi-disciplinary field, involving everything from applied mathematics to statistics and artificial intelligence to machine learning. And it's growing. This is because of advances in computer technology and processing speed, the relatively low cost to store data, and the massive availability of data from the Internet and other sources such as global financial markets.

For data science to happen, of course, you need data scientists. Because data science is so wide in scope, being a data scientist covers a range of professions. These include statisticians, operations researchers, engineers, computer scientists, actuaries, physicists and machine learners.

This variety isn't necessarily a bad thing. From my own practical experience, I quickly learnt that when solving data science problems, you need a range of people. Some can work in depth on theory and others can explore the application area.

But how should these data scientists be trained so they're prepared for the big data challenges that lie ahead?

Data scientists typically use innovative mathematical techniques from their own subfields to try and solve problems in a particular application area. The application areas – finance, health, agriculture and astronomy are just some examples – are very different. This means that each poses different problems, and so data scientists need knowledge about the

particular application area.

For example, consider astrophysics and the [Square Kilometre Array](#) being built on the southern tip of Africa. It will be the world's largest radio telescope when completed in the mid-2020s. The array of telescopes is said to receive data at one terabyte per second and researchers are typically interested in analysing the masses of data in order to detect tiny signals engulfed in white noise.

In finance, researchers exploit large data bases very differently: for example to learn more about their customers' credit behaviour.

The most established subfields of data science are statistics and operations research and it might be worthwhile to learn from the established training programmes in these fields. Are universities training enough graduates in these fields? And is that training good enough?

Although students in these fields are well trained academically, many graduates in statistics and operations research lack knowledge about the fields in which they are expected to apply the mathematical techniques. They also tend to battle with real-world problem solving abilities, as well as lacking numerical programming and data handling skills. This is because those skills are [not addressed adequately in many curricula](#).

So, drawing from these failings and the lessons of established data science subfields, what should universities be teaching aspiring [data scientists](#)? Here is my list.

- Mathematical and computational sciences, including courses in statistical and probability theory, artificial intelligence, machine learning, [operations research](#), and computer science.
- Programming skills;
- Data management skills;

- Subject matter knowledge in selected fields of application; and
- Professional problem-solving skills.

This list could be expanded at the postgraduate level. And, whether at undergraduate or postgraduate level, all of these courses should have a practical element. This allows students to develop both professionalism and problem-solving skills.

For instance, at the Centre for Business Mathematics and Informatics at South Africa's North-West University, my colleagues and I have organised a professional training programme that sees students working for six months at a client company to solve a specific industry problem. These problems are mainly in the financial field; for example, models to predict a customer's ability and willingness to pay, models for improving collections and models for fraud identification.

This helps students to develop the necessary skills to function in the working world, handling real data and applying it to real problems rather than just working at a theoretical level. It also, as a colleague and I have argued in previous research, helps to close the academia-industry gap and so makes data science more relevant. The BMI programmes have been recognised and commended by international experts.

Data science, as a field, is only going to grow over the coming decades. It is imperative that universities train graduates who can handle enormous tranches of data, work closely with the industries that produce and apply this data – and make data something that can change the world for the better.

This article is republished from The Conversation under a Creative Commons license. Read the original article.