techworld FROM IDG Insights

Issue: **January 2018**

THE COMPLETE GUIDE TO: **DATA SCIENCE**

Credit: iStock

Produced by IDG

Credit: iStock

Expert analysts

Data scientists. Every enterprise wants more of them, the biggest tech companies in the world are throwing money at them straight out of university, and demand shows no sign of tapering off.

As companies generate ever-greater volumes of business, operations and IoT data, the demand for people that can make sense of that information and present it back to the business to improve efficiency, and even create whole new revenue streams, is naturally appealing to a modern day enterprise. This is especially important with the rising interest in machine-learning algorithms and the value they can deliver to a business, so data scientists with experience running machine learning projects come at even more of a premium.

Here we run through what a data scientist can bring to an organization, what tools they are using, how to hire them and some examples of the results a good data science team can bring to real-life enterprises such as GE Digital and the NHS. Scott Carey

Contents

4 How to get a job as a data scientist

10 How to use data scientists in the enterprise

17 Best data science tools

22 NHS uses data science to cut down on A&E visits

25 Data science helping GE Digital drive industrial IoT

29 Plugging the data skills gap

Credit: iStock

How to get a job as a data scientist

Here we look at what skills you will need, what qualifications are expected, and some practical tips to achieving a successful career

With the rise of big data comes the need for more highly skilled people to mine and interpret that data for businesses. This is the role of a data scientist, the job that *Harvard Business Review* called “the sexiest job of the 21st century” back in 2012.

With more and more tech companies looking to make sense of their customer data, and with salaries topping out at £100,000, you can see why graduates with quantitative degrees – think mathematics, computer science, or astrophysics – are becoming data scientists.

What is a data scientist?

A data scientist's role is to derive actionable insights from huge data sets. This is different to a data engineer, whose primary role is to store and prepare that data, so someone with expertise setting up and maintaining large databases. The skills required of a data engineer tend to be more technical, with knowledge of Hadoop, SQL and NoSQL databases.

"Data engineers build massive reservoirs for big data," explains Sophie Adelman, head of sales EMEA for Hired.com. "They develop, construct, test and maintain architectures such as databases and large-scale data processing systems. Once continuous pipelines are installed to – and from – these huge 'pools' of filtered information, data scientists can pull relevant data sets for their analyses."

Qualifications and skills

Adelman says a strong undergraduate degree in 'quantitive' subjects such as mathematics, economics, finance or statistics is key. "You do see a lot of PhD and Masters coming out, as data science is a good way to apply what they have learned," she reveals.

In terms of hard skills, a data scientist will be expected to know how to interact with, and query, a database, so knowledge of analysis and data modelling skills such as Apache Hive and Pig and programming languages Python and R are useful.

However, Adelman adds it is the soft skills that many candidates fall down on. “Technical skills are important but people tend to over emphasise that. Commercial experience and soft skills are equally important,” she says.

“If people don’t have the ability to understand what the problem they are trying to solve is, and communicate it in a way that others can understand, it is difficult to be a great data scientist.”

Her advice? Go in search of more applied experience if you are a student or recent graduate. “You could use Kaggle to get more real-world experience and learn new techniques, or work as a financial analyst in banking, or take Coursera courses, or take on research projects.”

Nuno Castro, director of data science at Expedia also sees business awareness as an important skill. “People who can understand the organization from a commercial perspective, and who create relevant relationships in the organization will be more successful,” he argues.

He also puts a premium on problem-solving skills. “You and your team will typically be set a high level objective for which you need to determine the best cause of action. Unlike other areas, there is no one recipe for data science. Often you are not trying to find the right answer to a question, you’re trying to find the right questions to ask in the first place.”

If you speak to enough data scientists one thing you will hear is how much time they spend cleaning up data rather than analysing it. Sandra Greiss, a data scientist at online retailer Asos, says that even though this takes up 80 percent of her time, and the availability of tools for data cleansing (Trifacta, OpenRefine, DataWrangler), she would only ever want to do it by hand.

"It is frustrating, but I think it is also a relief when you are done with it as you will be using something which is correct," she says. "I don't think you would want to rely on a tool. You have to see it yourself."

Machine learning

One skill that is of growing importance in data science circles, and within the enterprise, is machine learning.

"Machine learning is a no-brainer to me. That is the true heart of data science," argues Mike Ferguson, an analyst at Intelligent Business Strategies.

"People want to have a pattern detection and a view into the future, so the traditional career in reporting is no longer enough, which is a key reason machine learning is critical. The days of taking data out of a database and doing the analysis somewhere else is done, the data is too big."

Asos' Greiss has seen it grow in importance within the industry: "I think it was already important, and I was asked about it at interview, but I don't have machine learning skills. I think machine learning is something you can pick up pretty quickly if you want to. Now it is something they will probably ask for because it has expanded so much and there is so much free online material available to give you some idea about how to do machine learning."

Employer perspective

Castro at Expedia says that a great data scientist must be "persistent, highly energetic and motivated". His advice for any prospective candidates is to: "Follow lots of other data scientists on social media – Twitter – read blogs, learn a new data science technology, practice on Kaggle or possibly enrol in an intensive data science course.

"After you've done that, try to get a data science internship. Make sure that you'll be working on a cool end-to-end data science project or a deep dive on a specific piece with a measurable output, e.g. a new algorithm that you can A/B test, rather than just doing what everyone else doesn't want to do, e.g. unit tests, though you will still learn."

From a technical perspective, Castro says it is better to learn using open-source technologies and skills, such as Java, Hadoop and Spark ML. This way, "when the next technology buzzword arrives you will be ready. If you spend your days working with a proprietary technology with its own programming language and workflow, how transferable is that and how will that add value to your CV?"

Gary Damiano, vice president of marketing at NoSQL database specialists Couchbase believes "the two things you need to consider in hiring a data scientist are: how are you going to use them and how does their skill set match the use?" For data scientists he wants to see "a programmer skill set backed by deep understanding of statistical regression metrics".

For data engineers, or "wrangling" as he calls it, they need to be "a database or spreadsheet ninja who can identify and build complicated data relationships and break them down into data segments that lend themselves to presentable views of relevant insights".

Data scientist perspective: Sandra Greiss

Asos data scientist Greiss's Linkedin profile summary is a single line: "I am not looking for a career move.

Thanks!" This is because she is approached by recruiters on the platform on a daily basis, often with badly-prepared pitches. "It is pretty frustrating, it's a shame that I feel like they don't all make the effort to know the candidate well," she says.

When she graduated from Warwick University with a PhD in astrophysics, Greiss turned down an offer of triple her entry salary at e-commerce company Lyst to go and work in the finance sector because "I wasn't interested in the topic or the project so I wouldn't give it my best".

Even working at a start-up like Lyst brought challenges though. "Finding a job wasn't as hard as I thought it would be, but adapting to the industry was challenging. The terms people use in companies and even the coding style was different, in academia they want performance and speed but aren't as particular."

Skills gap

If you can master the skills of data science though the rewards are plentiful. Hired's 2016 *Mind The Gap* report shows that data scientists are being paid increasingly well as employers look to attract them away from traditional roles in financial services. In the 18 months leading up to that report, salary offers for data scientists had risen by 29 percent. The only role that has increased more is security engineer, at 31 percent.

More recently, Hired's *State of Global Tech Salaries Report 2017* pegged the average UK data scientist salary at £56,000 for 2016, which is up 5.2 percent year-on-year. Scott Carey

Credit: iStock

How to use data scientists in the enterprise

Yandex Data Factory experts explain how to introduce data science into a business

Machine learning has become a buzzword in business technology but the implications of applying it are often overlooked.

“The major problem is that data science is science itself, and businesses aren’t very well accustomed to using scientific methods of decision making,” explains Jane Zavalishina, CEO of machine learning and data analytics specialists Yandex Data Factory.

The company emerged as a spin-out from multinational technology corporation Yandex, the operator of the largest search engine in Russia. In December 2014, the firm extended the capacity in data science it developed to support this core product into providing machine learning based services for industry applications by launching the Yandex Data Factory.

The Yandex Data Factory team establishes its findings through a process of experimentation, and its success can only be judged once the experiment concludes.

“When you delegate some work to your employee, ideally you expect more or less a complete level of results,” says Zavalishina, “but it works differently with data scientists, because with data science you cannot expect guaranteed results.”

Failure will be a legitimate outcome of any data science project and this is a prospect business managers must accept.

What makes a data scientist tick?

Working with data scientists requires an alternative approach to business in which logic overrules creativity and reality trumps belief. In other words, it depends on fact and logic rather than imagining what could be possible.

It will be a struggle, then, to task data scientists with questions that they fundamentally consider meaningless.

“It sounds like division by zero, it doesn’t make sense,” argues Zavalishina. “The problem is you can’t make them do this; you cannot motivate people to divide by zero. They start thinking you’re probably an idiot, which doesn’t make your work with them better.”

They need to understand the project and believe that it makes sense. If they are approached to use machine learning to improve systems, for example, they will need enough data to measure meaningful results.

"A lot of decisions in business are made by intuition, that's why there is no need to measure everything in regular business," reveals Yandex Data Factory COO Alexander Khaytin. "But then when it comes to a data science project or to communication with data scientists you can't just tell them, 'do this stuff, I feel it's going to be good.' It doesn't work."

Asking the right questions

Predictive analytics modelling relies on algorithms that tend to be far more complex than more traditional statistical systems. They can be difficult to explain.

The retail industry often uses data science to better predict stock replenishment requirements for weekly item orders. The results can amaze, but there are so many factors to take in that the process itself is often hard to communicate.

"It's just impossible to explain to someone who cannot grasp data complexity, but because it cannot be explained, you cannot decide how good it is just based on your common sense or business intelligence," says Zavalishina. "You need to make sure that you know what it is you want to improve, and how you measure results. It's not creative. It's specifics and what it tried to predict or optimize. It's like dealing with mathematicians. You ask the question and then you will receive exactly the answer to this question."

If your question is wrong don't expect the right answer. It's a surprisingly common problem, as companies often lack thorough planning on their objectives and the measurement of assessing them.

"We were working with this big retail company and they asked us to build a model which would predict how much of each and every item will sell the next week," Zavalishina recalls. "We tried it with one item, but the problem was they realised that [the prediction] is practically no use for them."

Their model was precise, but the company was ordering its product in packages of six rather than as individual items. If the prediction called for seven items next week, they would need to answer a different question. Should they buy one or two? It may appear a small change, but it meant they had started at the wrong place. The model became entirely different, because the parameters for optimization had shifted.

Data science requires careful planning. The company received the right answer, but should have asked a different question.

Failure on the route to success

The optimization model provided to another retailer suggested that the expensive and unusual products they rarely sold weren't worth ordering at all. The decision was mathematically logical, but that doesn't mean it made business sense. Such items can be crucial to the shop's identity and customer base.

"You are pretty much guaranteed that with your first data science project or machine learning project you will need to get back and rethink what the metrics are and what the goals are," argues Zavalishina.

Yandex usually recommends customers begin with projects that are very specific and short, to avoid the risk of a long-term investment in a project that could have meaningless results. This method allows companies to make piece-by-piece improvements across the board.

Another company had their own system to determine which customers were sent certain offers. Yandex would use the recommendations of a statistical model produced by a machine learning algorithm to determine how a random slice of the customer base was contacted. The rest of the customers were contacted according to the previous system, and the company then compared the conversion rates of offers into sales.

The only problem was that the offers were sent to the control group on Friday and to the experimental group on the weekend. The diverse patterns of behaviour at the different times of contact made any comparison meaningless.

Business managers often ask Yandex whether they should take courses in machine learning or data science to understand how the technology could benefit their organizations.

“What we usually answer is actually no, it doesn’t make any sense,” reveals Zavalishina. “It won’t make you data scientists, so it won’t really help you. If you want to be able to apply the technology in your work, you are much better off learning the scientific method and measuring and experimentation. Basically, we need a more scientific approach in the business if you want this technology to bring results.”

Accepting uncertainty

Businesses need to embrace the scientific culture. Negative results don’t mean the work has failed, they only prove that the optimization didn’t work.

The responsibility within the corporate structure is another challenge. Yandex was once approached by a client hoping to optimize its advertising spending. The algorithm developed promised the same level of response while saving 20 percent in costs.

Implementing the results proved more challenging than attaining them. The staff responsible for this project were paid bonuses based on their plans and decisions behind what they should buy to achieve optimal results.

“So now they have this model, which provides them with recommendations, and mathematically it is proven that the recommendations were better, but the problem is their responsibility,” explains Zavalishina.

Data science projects acknowledge the different responsibilities and priorities that can exist in the same business. This team was expected to implement a model that could result in cuts to their bonuses.

“When it comes to a scientific approach it’s much more rational, much more measurable, and this can be quite a conflicted situation,” adds Khaytin.

“The usual decision-making purpose is going to be at least disrupted. For example, an expert can tell you ‘I have an intuition, I have an idea, it’s going to be that way’. In our hands you have some data science tool, some data science project and it’s totally different, there is no intuition, there is no place for it.”

Integrating business and scientific approaches is a complicated process that requires patience and understanding. Yandex also worked with a steel manufacturer on optimizing the balance behind the mixture of materials used in the production process. The quality was improved by increasing the quantity of a certain substance, but the more of this substance that was added the more expensive the production. Yandex used historical data to make an accurate model of how best to balance the quality and cost of the mixture, returning with a recipe provided by a machine-learning algorithm.

"This recipe often doesn't make sense to them," says Zavalishina. "They look and say 'no it won't work, I cannot do that, I'm not accepting this, I'm doing something different'.

"The funny thing is it will bring better optimization, but on the other hand you have the experts' [preferences], so how do you deal with that? They are basically not using 80 percent of your recommendations.

"We came up with a solution, which would be another algorithm which looks at the recipe we provided, and on top of that builds the prediction of how probable it is to be accepted by that operator. So we optimized the recipe so they became a bit a less optimal from a strictly mathematical point of view, but much more probable to be accepted by humans."

Fears have long been expressed that artificial intelligence could destroy humankind, but the marriage between man and machine learning remains at the foundation of data science. Tom Macaulay

```
irror_mod = modifier_ob.
et mirror object to mirror
irror_mod.mirror_object =
peration == "MIRROR_X":
irror_mod.use_x = True
irror_mod.use_y = False
irror_mod.use_z = False
_operation == "MIRROR_Y":
irror_mod.use_x = False
irror_mod.use_y = True
irror_mod.use_z = False
_operation == "MIRROR_Z":
irror_mod.use_x = False
irror_mod.use_y = False
irror_mod.use_z = True

election at the end -add
_ob.select= 1
er_ob.select=1
ntext.scene.objects.active
("Selected" + str(modifier
irror_ob.select = 0
bpy.context.selected_obj
ata.objects[one.name].se

int("please select exactly

- OPERATOR CLASSES ----

types.Operator):
X mirror to the selected
ject.mirror_mirror_x"
ror X"

ontext):
ext.active_object is not
```

Credit: iStock

Best data science tools

Some of the best data science platforms available for modelling and deploying machine learning and advanced analytics

Platforms that allow data scientists to build and deploy algorithms are increasingly important as businesses look to operationalize their data faster than ever before.

Gartner defines data science platforms simply as “engines for creating machine learning solutions”. For the sake of this article we have broadened Gartner’s definition to include everything from data science workbenches, where teams can collaborate on code and deploy it themselves, to guided data science solutions.

It is important to remember that all data science platforms are relatively immature and none are a silver bullet. “Data science is not plug and play,” Matt Jones, lead analytics strategist at Tessella Analytics told us. “Platforms are fine, but they need to be trained by someone who understands the data and the context it exists in. If you’re outsourcing data science to a tech vendor, be absolutely sure they understand your business and your data.” With that in mind, here are some of the best and most popular data science platforms, from open source to established vendors, being used by enterprises today. Our top picks are:

- Microsoft Azure machine learning platform
- Domino Data Lab
- Cloudera Data Science Workbench

1. Microsoft Azure Machine Learning

Microsoft provides data scientists with a fully managed cloud service for building and deploying predictive analytics into live environments with its Azure Machine Learning platform. It comes with built-in packages to support custom code in your preferred language, be it Python or R, and a plethora of documentation for data scientists to get started.

The Azure platform lets data scientists deploy models into production quickly as a web service and then share them on the Azure marketplace to gain exposure. Customers include Carnival Cruises, JLL and Fujitsu.

2. Domino Data Lab

California-based start-up Domino Data Lab’s platform is another ‘workbench’ solution, allowing data science teams to do modelling on their preferred data sources, using whatever tools and programming languages they are comfortable with and to collaborate and deploy models straight from Domino as APIs.

It then acts as a hub for all data science activity, elastically provisioning compute in the cloud and deploying in a consistent, secure manner so that IT can take a back seat. Data science teams at insurers Zurich and Allstate are both customers of Domino.

3. Cloudera Data Science Workbench

Analytics vendor Cloudera launched its ‘Data Science Workbench’ in March 2017 following the acquisition of Sense.io in 2016. It is intended to be a platform where data science teams can work with their data in popular programming languages such as R, Python and Spark in a secured-by-default, collaborative environment. The idea is to make the modelling and deployment of machine learning and advanced analytics within the enterprise at far greater speeds than if they had to worry about anything other than the actual data science.

4. SAS Viya

Analytics and BI vendor SAS provides data science and machine learning capabilities through its Viya platform. This is an example of an analytics vendor providing customers with a platform where they can take their advanced analytics work out of self-contained clusters and into an environment where they can be deployed in a secure, consistent way. “We try to enable people to use what they want to use, but not reinvent the wheel every time,” Peter Pugh-Jones, head of technology at SAS UK and Ireland told us.

5. Dataiku

The French start-up Dataiku provides a host of guided data science and machine learning processes on its platform DSS. The platform has a level of abstraction so that anyone using it can either code in Python, Pig, R, Hive, and so on, or use drag and drop functionality to wrangle and model data. The platform allows teams of data scientists, data analysts, and engineers to prototype, build and deliver data solutions into the businesses from a single place. Previous customers include L'Oreal, Trainline and AXA insurance. In its more recent releases Dataiku has added point-and-click capabilities (called 'visual recipes') for data preparation, the ability to monitor model performance during training, and support for Python 3 with a new code editor.

6. IBM Data Science Experience

IBM offers a range of data science tools and is preparing to release an IBM Watson-guided machine learning platform. The current iteration comes with built-in learning, so that data scientists can improve the more they engage with the platform, collaboration features and notebook tools for working with popular programming languages, such as Jupiter Notebooks for Python and RStudio for R. The enterprise version of the platform retails at $9,200 per instance per month and provides managed Spark clusters and flexible storage.

7. RapidMiner

Open-source data science platform RapidMiner helps the likes of BMW, Samsung, and Barclays launch data science projects. Tools on the platform include Studio, for visual data science workflows, Server for operationalizing models, and Radoop for workflows using Hadoop data. For larger customers or projects there are enterprise versions of the platform which range from $2,500 to $10,000 a year depending on the rows of data.

8. Knime Analytics Platform

The open source and free Knime Analytics Platform looks to give data scientists a blank canvas to work on projects using various data sources and the tools they are comfortable with in a scalable environment.

The open platform comes with thousands of native nodes and modules, extensive documentation and pre-packaged advanced algorithms to get started quickly. Data scientists can toggle between single computer, streaming or big data on top of or alongside existing infrastructure and makes sure that everything is backwards compatible and easily portable for flexibility.

9. Splunk Machine Learning Toolkit

Big data specialist Splunk has moved into more integrated machine learning within its platform over the past year or so, but the vendor also provides a Machine Learning Toolkit for custom models. The advantage of using Splunk over other workbench solutions is that you can model straight on top of machine-generated data – Splunk's area of expertise – so security and IoT use cases are a natural fit. The Toolkit is a guided workbench for data scientists to model and deploy algorithms in the most popular programming languages. There is also a library of pre-built Python algorithms for popular use cases, and plenty of documentation and tutorials to get started straightaway. Scott Carey

Credit: iStock

NHS uses data science to cut down on A&E visits

The NHS is deploying its data scientists to analyse its 111 emergency call logs and predict attendance more accurately

NHS Digital hopes to ease the burden on local accident and emergency (A&E) units by analysing call data into its 111 emergency telephone service.

Speaking at the Tech for Britain conference in London, Daniel Ray, director of data science at NHS Digital said: “Where we have started to generate new big data and apply data science is in areas like NHS

Pathways. So if you phone the 111 service, up until recently no one had analysed that data."

The NHS received 15 million 111 calls per month, two million of which end up with a patient being referred to a local A&E. The NHS 111 telephone helpline has actually increased the number of people turning up at emergency departments and calling ambulances since being launched, the opposite of its intent to ease the burden on A&E units.

Ray explained that NHS Digital has always collected 111 call data as "when a call comes in the handler sits in front of a clinical system and they capture information. What sits behind that is a clinical algorithm decision tree."

He noticed that the data from these calls was just "sitting there" when he arrived at the organization in 2016, so he kicked off the project to try to get more insight into the referral process and to hopefully improve the clinical algorithm to reduce the number of people being sent to A&E where possible.

With a small team of two data scientists, the organization started looking at what the patient actually does after a 111 call and secondly, if they should have referred that patient to A&E at all. By linking up the 111 call data and local A&E inpatient records he believed "we could potentially stop hundreds of A&E attendances simply by tweaking the clinical algorithm that the nurse fills out at the other end of the phone".

Predicting A&E demand

Ray, who joined NHS Digital as part of the new Centre of Excellence for Big Data and Data Science, is tasked with finding new ways to use data science to deliver better care across the health service. Another project he is working on is to try and improve the predictions the NHS can make around A&E demand.

By analysing streaming data from sources such as sporting event databases and the Met Office for weather, Ray claims to be able to increase its demand prediction models for A&E attendance by up to 25 percent. This is "enough for someone who is in charge of a hospital to redirect how many resource they need to cope", he said.

Something as innocuous as weather data is significant because "the temperature outside makes a massive difference to both the total volume of A&E attendance and the type of patients that come through the door", Ray revealed. "When the sun is out everyone is out breaking themselves." **Scott Carey**

Credit: iStock

Data science helping GE Digital drive industrial IoT

The General Electric subsidiary predicts that the industrial Internet will be worth $225 billion by 2020

GE Digital is betting big on the power of data science and the Internet of Things (IoT) to transform industries of every sector.

In July 2017, the General Electric subsidiary hosted a data science event at its digital foundry in Paris that brought together data scientists, academics and industry representatives as a statement of intent in its drive towards the digital industrial era.

"The economic weight of the industrial Internet is as big as the consumer Internet, but not everybody knows this, and you really see that the academic community is starting to understand this," reveals Vincent Champain, GM of the European Foundry of GE Digital.

"Computer science is now connecting with industry and not in a specific case, but really becoming mainstream, because now with cheap sensors [and] cheap computation, people have been able to gather huge stacks of data and the key now is to find the needles of value in those stacks of data."

Champain says that every company and sector invested in industrial data innovation is exploring a combination of advanced data science and the power of cloud computing with agile methods to roll out apps at high speed and low cost.

First though, he says it is important to note that the type of data scientist required for industrial applications differs from those traditionally working in the consumer space. "It's the real person mastering the law of physics around how assets behave," explains Champain. "Compared to the consumer space, physics plays a huge role and knowing the machine, knowing the physics, thermodynamics, and the chemical reaction is key to bring performance."

How IoT is transforming the industrial world

Investors and the media tend to focus on the more domestic, consumer Internet of Things, but as machinery becomes intelligent the spending in that segment could soon be dwarfed by that of Industrial IoT (IIoT). Accenture estimates that it could add $14.2 trillion (£10.9 trillion) to the global economy by 2030.

IIoT is also known as the Industrial Internet, a term coined by General Electric (GE) in 2012. The American industrial conglomerate claims that this market could be worth $225 billion (£172 billion) by 2020, by providing customers with the capability to autonomously monitor processes and make real-time adjustments.

GE has backed up its bold predictions with actions. In 2015 it founded GE Digital to drive its Industrial Internet business and released a cloud-based IoT operating platform called Predix that customers can use to develop apps that drive efficiencies by analysing real-time operational data.

Companies will spend €250 billion on IoT in 2020, according to research by Boston Consulting Group. GE Digital has developed a number of data-driven applications that will be competing for this investment.

The Predix platform lets customers analyse the productivity of equipment and drive efficiencies through maximizing energy use and revealing additional available capacity and potential defects, such as corrosion in industrial pipe systems. “You can send robots who take kilometres and kilometres of pictures of pipelines and then this artificial intelligence will tell you precisely at which kilometre, metre and centimetre you potentially have a problem,” says Champain. “That’s huge. Huge impact and it can really avoid catastrophes.”

Other aspects of industry that can benefit from IIoT optimization include inventory and supply chain management, remote monitoring of utilities with sensors, predicting equipment break downs, demand response, and tracking product orders and vehicles.

GE isn’t the only company tackling this area, though. Microsoft is working with rival airline engine maker

Rolls-Royce to predict when engines will require maintenance, using its Azure suite of cloud tools to monitor engine health.

GE plans of playing a leading role in IIoT

GE is also using the technology to transform its own business. In 2016, the company made $730 million (£559 million) of productivity savings through apps that range from utilizing machine learning to understand signals from equipment, to cutting the amount of scrap material generated in the tube cutting process in half.

The company and IIoT could also benefit from having political support in high places. The GE Digital centre was inaugurated in 2016 by a young minister of economy called Emmanuel Macron.

“He grew bigger, and we did too,” said Champain. “He’s really pushing this and he really understands what’s happening in the service space.” Tom Macaulay

Credit: iStock

Plugging the data skills gap

Academia and big business are the perfect match when it comes to solving data science problems, but who benefits the most?

With the rise of big data and the lack of data scientists in the market, businesses are reaching out to academia more and more to help solve some of their thorniest technical problems.

In theory, it is a match made in heaven. Academics want interesting problems to solve and businesses have plenty of them when it comes to making use of their new-found reams of big data. In a world where data scientists are the scarcest of resources, partnerships between universities and businesses are helping to plug the data science skills gap.

Alice Jacques, senior data scientist, consumer insight at Channel 4 summed it up during her talk at the recent DataIQ conference in London. “Businesses have data and real problems, academics have experts and teaching capacity,” she explained.

The benefits for both sides are clear to see: industry gets cheap access to data science talent and universities can tout industry experience to attract students, boost funding and help with their rankings.

Professor Patrick Wolfe from University College London (UCL) told us that he sees universities moving away from the “old-fashioned ivory tower model” as they look to engage with society more. He believes that by linking arms with big business, universities have an opportunity to “apply new ideas immediately”.

“If I wanted to work on a network 30 years ago there wouldn’t have been much data to work on,” he added. “What has happened now is there is a natural bridge.”

Channel 4

Jacques told the conference that the media company worked with two UCL PhD and four master’s students on a soon-to-be released Netflix-style recommendation engine for its on-demand service All4.

She said that the current batch of PhD students “are doing their PhDs in recommendation systems and computational statistics, stochastic modelling and time series,” which are all skills that have helped Channel 4 develop its recommendation system. “We have six mini-experts that understand a single problem very well.”

Jacques urged data scientists to get out to academic conferences more. “They scare you and remind you that academics didn’t stop when you left [university],” she argued. “They expose you to new ideas, which you can try to fit into your business problems. It is speed dating for data science ideas.”

Poaching talent

This trend goes further than collaboration and into outright talent poaching.

“The other side is a trend of academic researchers in areas like computer science sometimes being poached away to industry,” Wolfe added. “If you are excited about running algorithms at large scale you can’t really do that sitting behind a university desk.”

According to Wolfe, industry experience simply makes his students “more immediately marketable”, but that demand for data science skills will continue to outstrip supply because “universities work on a slightly slower scale” than industry.

On the other hand, universities use these relationships as a way to attract master’s students. For example, the data science school at City University of London states: “When it comes to the big data and data science area, we have established relationships with organizations including the BBC, Microsoft and The British Library, so you can be confident that with City, your access to professional experience is unparalleled.”

The University of Lancaster’s data science institute website lists nearly 40 industry partners that students can access. The website looks to attract new partners by promising “access [to] skills in data mining, programming, statistical modelling, statistical inference” and the ability to “gain an early view of talented data science professionals”.

Conclusion

It certainly looks like closer relations between academia and business can bring huge benefits for universities in terms of funding future research and for students looking to boost their job prospects.

The only concern will be if the relationship gets too cosy and academics lose their all-important scope to come at problems in a broader sense than simply solving a business problem. Scott Carey

IDG