Monday, 24 October 2016

Data scientists will be crucial in 'cognitive' tech economy

by Richard Waters (Financial Times, July 2016)

For a field supposedly starved of talent, data science seems to have been minting a lot of new experts in a hurry. The depth of interest was on display this week in San Francisco, where 1,600 people turned up for a data science summit organised by Turi, a company run by University of Washington machine learning professor Carlos Guestrin. Mr Guestrin argues that all software applications will need inbuilt intelligence within five years, making data scientists — people trained to analyse large bodies of information — key workers in this emerging "cognitive" tech economy.

Whether or not he is right about the coming ubiquity, there is already a core of critical applications that depend on machine learning, led by recommendation programmes, fraud detection systems, forecasting tools and applications for predicting customer behaviour. The adaptation of what was until recently the preserve of research scientists into production-grade business applications could point to a profound change in corporate competitiveness. The companies showing off their skills in data science and machine learning at the Turi event - including Uber, Pinterest and Quora — were all born in the digital era.

Some companies that grew up in the analogue world, like Walmart, are also investing massively in this field, says Anthony Goldbloom, the chief executive of Kaggle, a company which runs online data science competitions. But he predicts that they are unlikely ever to catch up with Amazon and its ilk, which have a head start and are moving fast. Repeated across different sectors, that could point to wholesale change in industry leadership as intelligent systems take a more central role.

One factor weighing on many traditional companies will be the high cost of mounting a serious machine-learning operation. Netflix is estimated to spend $150m a year on a single application - its film recommendation system - and the total bill is probably four times that once all its uses of the technology are taken into account, says a person familiar with its applications.

Many companies that were born digital - particularly internet outfits that have a deluge of real-time customer interactions to mine - are all-in when it comes to data science. Pinterest, for instance, maintains more than 100 machine learning models that could be applied to different classes of problem, and it constantly fields requests from managers eager to use this resource to tackle their business problems, says Jure Leskovec, its chief scientist.

Another problem for many non-technology companies is talent. Despite the surging ranks of data scientists, some skills are in very short supply, particularly in deep

learning - the most advanced form of machine learning. Of the freelance computer science experts who use Kaggle, only around 1,000 have deep learning skills, compared to 100,000 who can apply other machine learning tech- niques, says Mr Goldbloom.

He adds that big companies are often reluctant to bend their pay scales to hire the top talent in this field, even if the algorithms developed by a single high-paid expert can have a disproportionately large effect on their business.

The biggest barrier to adapting to the coming era of data science at "smart" appli- cations, however, is likely to be cultural. Some companies, like General Electric, have been building their own Silicon Valley presence to attract and develop the digital skills they will need. But they will have to push their new data scientists and machine learning experts out into operating divisions and bring them closer to line managers to reap the full benefits.

This confluence, between the science and the business practice is critical. It has become a truism to say that all managers will need to let their decision-making be led by the data from now on. But that requires a complete change in mindset that is easier said than done.

The challenge is made even harder, says Mr Goldbloom, by the fact that managers are required to redesign their work processes around the new "smart" applications, in ways that effectively design themselves partly out of a job. Despite the obstacles, some may master this difficult transition. But companies that were built, from the begin- ning, with data science and machine learning at their core, are likely to represent serious competition.