

Keynote Speech: Data Science and Machine Learning in Practice

7th Annual SKBI Conference 2017: Advances in Data Science and Implications for Business

Friday, 26 May 2017

Introduction

Good morning ladies and gentlemen.

It is my pleasure to be invited here to give this keynote speech about a topic I happen to be particularly passionate about.

Today's conference is also the two-month anniversary of the formation of the Data Analytics Group, or DAG for short. Hence a timely opportunity for us to share, from the onset, our approach – the direction in which we are heading – and what it means to us, in putting data science and machine learning in practice.

We will take a broader perspective on this topic.

Over the past two months, our team has been doing quite a fair bit of information gathering and exploration:

- We have been working with the various departments in MAS to see how we can enhance the way we carry out our own functions – whether financial supervision or central banking or industry development – through the smarter use of data and analytical tools.
- We have also been speaking to various organisations and individuals outside MAS – academics, financial institutions, other government agencies, service and software providers, start-ups, and others – to explore areas of collaboration, pilot testing, information sharing and partnership in the use of data.

We are now starting to put in place the necessary tools, infrastructure and skillsets to harness the power of data science to unlock insights, sharpen surveillance of risks, enhance regulatory compliance and transform the way we do work. Alan Turing first introduced the concept of the Turing Test in his 1950 paper – If you recall the Turing Test is a test in which a person needs to distinguish whether he is talking to a machine or person. In this 1950 paper, Turing states, “We can only see a short distance ahead, but we can see plenty there that needs to be done.”

Fast forward to 2017 and we are still grappling with a concept written in a paper in 1950, so you see, we have our work cut out for us.

We will share with you some of our thoughts on how we plan to achieve these goals and position MAS and the financial sector for the digital economy of the future.

Given the plethora of terminologies, especially in this field, I believe it is important to contextualise upfront the meaning of terms used. I want to first define what we mean by “data science”, and this is not intended as an absolute definition.

To us, “data science” is the business of algorithmically deriving insight from data.

Now, this is not the same as AI, or artificial intelligence – the insights and data themselves are inherently not “intelligent”. What we are trying to do is to obtain data-driven insights in a smarter and more systematic way.

We see four components in this endeavour:

- First and foremost – as one would expect - is getting the **data** in order.

This involves not only an understanding of the data we have and need, but also the necessary processes surrounding the data.

- Second, deploying the appropriate **tools**, or technological capabilities.

This includes machine learning, natural language processing and network analysis algorithms, amongst others, to find out what the data is saying or could say about patterns and behaviour.

- Third, putting in place the necessary **infrastructure** to facilitate the two components I have just mentioned. This entails making decisions on using:

- 1) Unified platforms for data collection and submission;
- 2) Cloud technology to run data analysis;
- 3) Reusable tools and code libraries to facilitate analysis; and
- 4) APIs for disseminating data and allowing scalability.

National Archives of Singapore

- Fourth, training people with the right and relevant **skillsets** in these areas,

so that we not only use the right algorithms on the right datasets with the right infrastructure, but understand what the algorithms are doing and what the results **mean** for decision-making. This goes beyond the science of data cleaning and understanding the math behind the algorithms; it is the **art of contextualising** the output of these algorithms.

For this reason, I personally believe that it is not enough just to be a data scientist, but to aspire to be a data artist.

We have now briefly introduced the four components to you – data, tools, infrastructure and skillsets. We see these four components as building blocks for any successful data science venture, building blocks that are *foundational*, because if any one of these components is missing or lacking, you cannot derive good data-driven insights. These four components are also *interconnected*, because the choices we make for any one of these components depends on our choices for the other components.

These form the baseline of our point of view. I would now like to focus the rest of my remarks on the practical implementation – in particular the approach we plan to take in putting data science into practice within MAS as well as in the

rest of the financial sector. We see two principle components in putting data science and machine learning into practice – one, *partnership* and two,

experimentation. We seek to adopt both.

Hub-and-Spokes Partnership

Within MAS, DAG works in partnership with the MAS departments, in what we like to call a “hub-and-spokes” model.

- First, while DAG may have some of the analytical and technical expertise in terms of machine learning, natural language processing and network analysis algorithms, it is the various departments in MAS who have the domain knowledge. As the hub, DAG is working with the respective departments as spokes, to understand the issues on the ground, the types of problem statements we want to solve, and then figure out the appropriate tools we need to answer those questions. Applying advanced algorithms blindly without understanding the context will get us nowhere.

This is true in any scenario – it is the fusion of analytical and domain expertise that will enable us to find new and more efficient ways of doing our jobs.

National Archives of Singapore

- Second, data science should not be a standalone concept – it should part and parcel of everyone’s operations and job. Whether now, or as a future aspiration and whether one is explicitly called a “data scientist” or not. To be operationalised, one needs to realise that the requirements of data science will vary from person to person – it is a **not one-size-fits-all situation**. So, it doesn’t make sense to think of data science as the domain of a single group.

We plan to work in a similar “hub-and-spokes” manner with the rest of the financial sector community, and we hope to partner with stakeholders in as many ways as we can.

The SupTech-RegTech nexus

In this day and age everyone seems to need their own term – and sparring none – SupTech (or Supervisory Technology) is our contribution. It is perhaps easier to explain by first touching on RegTech or Regulatory Technology, which is the application of technology to enhance the way regulatory compliance is done in financial institutions. SupTech is essentially the other side of the same coin – applying technology to enhance the way MAS carries out its financial supervision functions.

As an example, a potential application of RegTech is in the area of anti-money laundering and countering financing of terrorism (AML/CFT for short) where high risk AML/CFT cases can be used to train a machine learning algorithm to identify potential future cases that financial institutions can investigate further. Identified drivers can potentially help us understand in greater detail the distribution of risk, behaviour, effectiveness of intervention, and misses or false-negatives.

The reason I said two sides of the same coin, because this is not a competition between MAS and the financial institutions. Quite the contrary.

Due to the potential for data science to be applied in the RegTech-SupTech area to benefit individual financial institutions and the financial system as a whole we have set up a dedicated Supervisory Technology, or SupTech, Office within DAG.

The approach of the SupTech Office is as the coin.

National Archives of Singapore

- The RegTech side – to promote data analytics capabilities within the financial industry to foster innovation and make regulatory compliance more efficient and effective.
- The SupTech side – to work with the supervisory departments to conduct data analyses and enhance our supervision of financial institutions.

We want to both sides to build upon each other so that risks in the financial system can be managed more effectively and efficiently, both individually and collectively.

To senior management in financial institutions, I urge you to go beyond approaching compliance functions as mere cost centres, and tap on them as rich sources of methodology and insight. Find ways to leverage on compliance data to look for linkages and patterns that can help you understand your customers and business activities better. Explore ways to replicate the techniques developed in the regulatory compliance space to enhance the functions of your other business activities. We want to work with you to share tools, and insights, so that we can each go back with ideas and suggestions of how to improve the way we do our work.

We actively want to erode any friction in collectively progressing in our data analytics capabilities – this perhaps most importantly includes working jointly in sandbox-like environments, when we don't have all the answers in advance. Following in spirit one of my favourite quotes "When in doubt, ask".

The SupTech-RegTech efforts I have mentioned so far cover tools and skillsets, two of the four key data science ingredients. Let me now touch on how we will address the other two, namely data and infrastructure.

Facilitating data flows

I don't think I need to belabour the importance of getting the data in order so to apply data science meaningfully. Yet dealing with data is often the most underappreciated, onerous and time consuming part of the data science process. What we want to do, what we must do, is make dealing with data as seamless and painless as possible. Given that data volume is only going to increase, we must address these core data challenges head on, and make it fun in the process.

We plan to do more, and make data more easily available for financial institutions and researchers to use for their own analysis. Of course with the right controls in place, since in some cases we may not be able to share the detailed data, but we can publish aggregates. We also want to give back insights that we draw from the data, to increase mindshare within the industry and foster an environment of knowledge sharing.

Within DAG, this is where the Data Governance and Architecture Office will focus its efforts on. It will oversee the governance and management of data at MAS. It will seek inputs from the financial industry to see how best we can enhance and streamline MAS' data collection and management process, and put in place the necessary infrastructure and platforms to achieve this goal. It will also work towards publishing more data and insights to the public.

Sharpening MAS' in-house capabilities

Finally, we want to walk the talk, lead by example and put all the key elements of data, tools, infrastructure and skillsets in place within MAS to drive a data culture. We will hold ourselves to the same standard as what we expect for the financial sector. The Specialist Analytics and Visualisation Office in DAG is charged with precisely this objective, to drive the analytical agenda across all of

MAS. As we build up our data culture and push forward our capabilities, we pose a friendly challenge to the financial institutions to do better than us.

National Archives of Singapore

MAS as matchmaker

This brings me to the final element of our “hub-and-spokes” engagement. As we are all in the same journey together in figuring out how best to adopt data science to enhance the way we work, we believe that the most efficient way to do this is by sharing ideas and expertise.

To facilitate this, MAS will play the role of matchmaker. Over the past two months, we have been reaching out not just to financial institutions, but also to other stakeholders such as academia and other government agencies. We would like to build a community of like-minded practitioners and researchers to facilitate cross-fertilisation and experimentation of data science ideas in the financial space.

National Archives of Singapore

People

This is a nice segue into my next set of reflections, which is on the role of people, not machines, in this era of data science.

There is a common refrain that as we advance in our adoption of data science, data scientists will take over all jobs. Alongside the compounded fear that machines will also take over jobs. It is a very real concern amongst people not only in the financial sector, but also in other industry sectors. I would like to explicitly bring this up, and look at how we can work together to address these concerns, on a continuous and ongoing basis.

Before I do that, let me relate a story from more than a hundred and fifty years ago, when the typewriter was invented in the 1860s and started to be sold commercially on the market for use in professional and business correspondence. Knowing how to type was a rare skill initially, and typewriting was considered a specialised technical skill limited to a handful. As its adoption became more widespread, everyone wanted to be a typist.

And now, typing a word document is considered a basic requirement of many jobs.

The point I want to make here is that technological advancement will happen whether we like it or not. Instead of resisting change, let us embrace it, and do what we can – and must – to be prepared for the changes that lie ahead. No doubt, technology and data science, machine learning, etc. will change the nature of current jobs, but that doesn't mean there will be no jobs. In fact, I believe there will be an abundance of jobs – some of which we have not yet imagined.

As I alluded to at the beginning of my speech, people will continue to play a very important role in putting data science into context. We will still need people to look at the output of complex machine learning algorithms, and interpret the results, decide what it actually means, and what implications there are for decision and policy making. To figure out what is the ultimate objective of the problem we are trying to solve, and what is the appropriate choice of algorithms and techniques we should use. Insights may be drawn from data that has been processed through the most complex of algorithms perhaps, but ultimately we as people need to put the analysis in context, draw meaning from it, and make the final decisions.

We are taking this transformation head on, and preparing ourselves and the financial industry for these changes. Within MAS, part of DAG's role is to develop data analytics training programmes for MAS staff, so that we will be well equipped to employ data science to enhance our work processes.

We are mirroring these re-skilling and up-skilling efforts in the rest of the financial industry. As we explore and push the research and development of technology to transform work processes in the financial sector, we will not ignore the need to assess the impact of the adoption of such technology on the workforce. We have begun, and will continue to engage with industry and academia on these issues, and call on engagement for partnerships.

In addition, we will be looking at how best to boost our initiatives to enhance the competencies of financial sector professionals. The Financial Sector

Development Fund, or FSDF, has several existing schemes to support the development and upgrading of skills and expertise required by the financial

services sector. To name a few:

National Archives of Singapore

- The Financial Scholarship Programme co-funds Singaporeans pursuing postgraduate studies in specialised finance areas.
- The SkillsFuture Study Award for the Financial Sector supports finance professionals attending programmes or courses to deepen their skills in areas particularly where industry expertise remain in short supply, including in data science.

There are many funding schemes in place, but we want to see where we can do more to ensure that the financial sector workforce continues to stay ahead of the curve in this digital transformation journey.

In addition, the Institute of Banking and Finance will be launching a mobile learning platform, which will support the dissemination of mobile learning content for areas such as digital awareness, data-driven decision-making,

human-centred design, agile thinking, future communication, and risk and governance in the digital world. This set of six core future-enabled skills will help

enhance financial sector workforce mobility and ability to respond to industry changes, as digitalisation takes on a more important and integral role in banking and financial processes.

In Conclusion

Isaac Asimov once said, “I do not fear computers. I fear the lack of them.” With how much we depend on computers and their algorithms today, I can’t imagine how life in the financial industry will be like if we were to go back to the era of typewriters to generate documents and punch cards to process data.

The adoption of advanced computing and data science in the financial sector will inevitably continue, and it will undoubtedly change the way we work and transform the way financial institutions operate.

We will work together with all of you – academia, start-ups, and industry practitioners – to build an environment which has the necessary data, tools, infrastructure and skillsets to facilitate the use of data to gain new insights, improve decision-making, and make Singapore’s financial industry well-placed for the digital economy of the future.

There are many moving parts, many necessary ingredients, many stakeholders, but none of us should walk the journey alone. It is our collective responsibility to seek out the opportunities in this revolution, and not only ride the wave of disruption, but to facilitate the transition of culture.

Let us embrace these changes with open hearts and minds, and work together in our data science journey to create value for the economy and financial sector and at the industry, corporate and individual levels.

Thank you.

National Archives of Singapore