South African Institute of Taxation

Big Data and Analytics

The impact on the accounting profession

An overview of the report of the ICAEW

YOUR KEY TO THE TAX COMMUNITY

Introduction

- Data is not new
  - Always used to support decision-making and manage operations
- So, what is so special about **big data**?
  - How do businesses generate value through it?
- The effective application of **analytical techniques** to **big data**
  - Personalised services, optimised operations, better risk management
- More focus on **hard evidence and facts**, and less of guesswork and assumptions
- Volume and breadth of opportunities to use data are transforming the business environment, creating opportunities for **disruptive business models**
- Hence the **three important questions** we're looking at today:
  - What is big data and what's creating it?
  - What can analytical techniques do, what are the opportunities and risks?
  - How can accountants exploit these trends?

Source: ICAEW Thought Leadership Tech Faculty

cpd campus

What is big data and what's creating it?

What is big data…?

- **Not just about volume**
- Also about complexity and speed
- Characterised by the "three V's"
  - Large **V**olumes (quantities/sources) of data
  - High **V**elocity (speed/pace/rate) of data flows
  - High **V**ariety of data (especially unstructured and semi-structured data, such as text and images)
- Has also played a key enabling role in increased machine learning, thereby supporting the development of artificial intelligence

Source: ICAEW Thought Leadership Tech Faculty

What propels big data?

- Three factors played most significant role to now
  - Growth in computing power
  - New sources of data
  - Infrastructure of knowledge creation
- Let's explore these…

Source: ICAEW Thought Leadership Tech Faculty

Computing power and storage

- Seen as the **core enabler** of big data developments
- The capture and processing of **entire** data sets, regardless of size and complexity
- Computing power grows '**exponentially**'
  - The Chinese chess board (eventually 18,446,744,073,709,551,616 rice grains)
  - Change starts small, but quickly becomes difficult to comprehend
  - Brynjolfsson and McAfee argue in their book, *The Second Machine Age*, that we find ourselves currently in the *second half* of the chessboard, with large improvements in short periods of time
- **Cloud computing** provides access to substantial computing resources, which they don't have to buy, but can use as and when they need to
- **Software advances** in processing and storing capability (new software supports handling of unstructured data sets better than traditional database management systems, e.g., Dext, GPT4, etc.)

Source: ICAEW Thought Leadership Tech Faculty

Data sources

- Increases in computing power in turn enable data collection and processing from many new sources, such as:
  - The internet (clickstream data from searches, sites visited, goods viewed, actual transactions)
  - Social media (status updates, comments, likes, photos, videos, networks of contacts)
  - Mobile technology (increased opportunity to create internet and social media data, location of individuals)
  - Open data (large amounts of primarily public sector data, such as geo-spatial, transport, government financial, public service data)
  - The IoT (i.e., embedding computer chips and sensors in physical assets such as buildings, machines, domestic appliances, clothes) which generate data
- Businesses increasingly use digital technology in sales and marketing, customer management, supply chain, internal communications, which all create data they can use
- Advances in the use of unstructured/semi-structured data, e.g., email, text, CCTV, pictures, and voice (what we talk about)
- Massive "datafication" of our activities, e.g., what we buy, look at but don't buy, travel routes, photos we take, where we go, who we visit

Source: ICAEW Thought Leadership Tech Faculty

Infrastructure

- Digital infrastructure enabled new types of collaboration and knowledge creation (e.g., crowdsourcing, open source software)
- Formed new communities and led to insights from unexpected places (data scientists can sometimes know nothing about the data, but spot patterns/relationships/correlations therein)
- Flexible nature of the digital environment makes all different kinds of collaboration, knowledge sharing, and creation possible
- *Example: Collaboration on automated language translators (word-for-word ☹; phrase-for-phrase 😐; neural network-based sentence-for-sentence translation ☺) = better joint understanding of problems to be solved, enabled by improved knowledge infrastructure*

Source: ICAEW Thought Leadership Tech Faculty

Analytical techniques and tools

- what can it do, and what are the opportunities and risks? -

"Big data analytics" – the buzz…

- Big data in itself, doesn't mean much…
- We need analytical techniques to help us make sense and gain insights from it
- Analytics and visualisation tools have become more user-friendly, i.e., less reliance on technical experts
- Costs have decreased, often linked to cloud applications (i.e., pay when using them)
- Sometimes mere ability to visualise large amounts of data is sufficient, using data visualisation software
  - Identify patterns, outliers and exceptions
- Linking of data from different sources together provides new insights
  - Advanced statistical analysis of data, using more variables, more real-time data, and more sophisticated data modelling

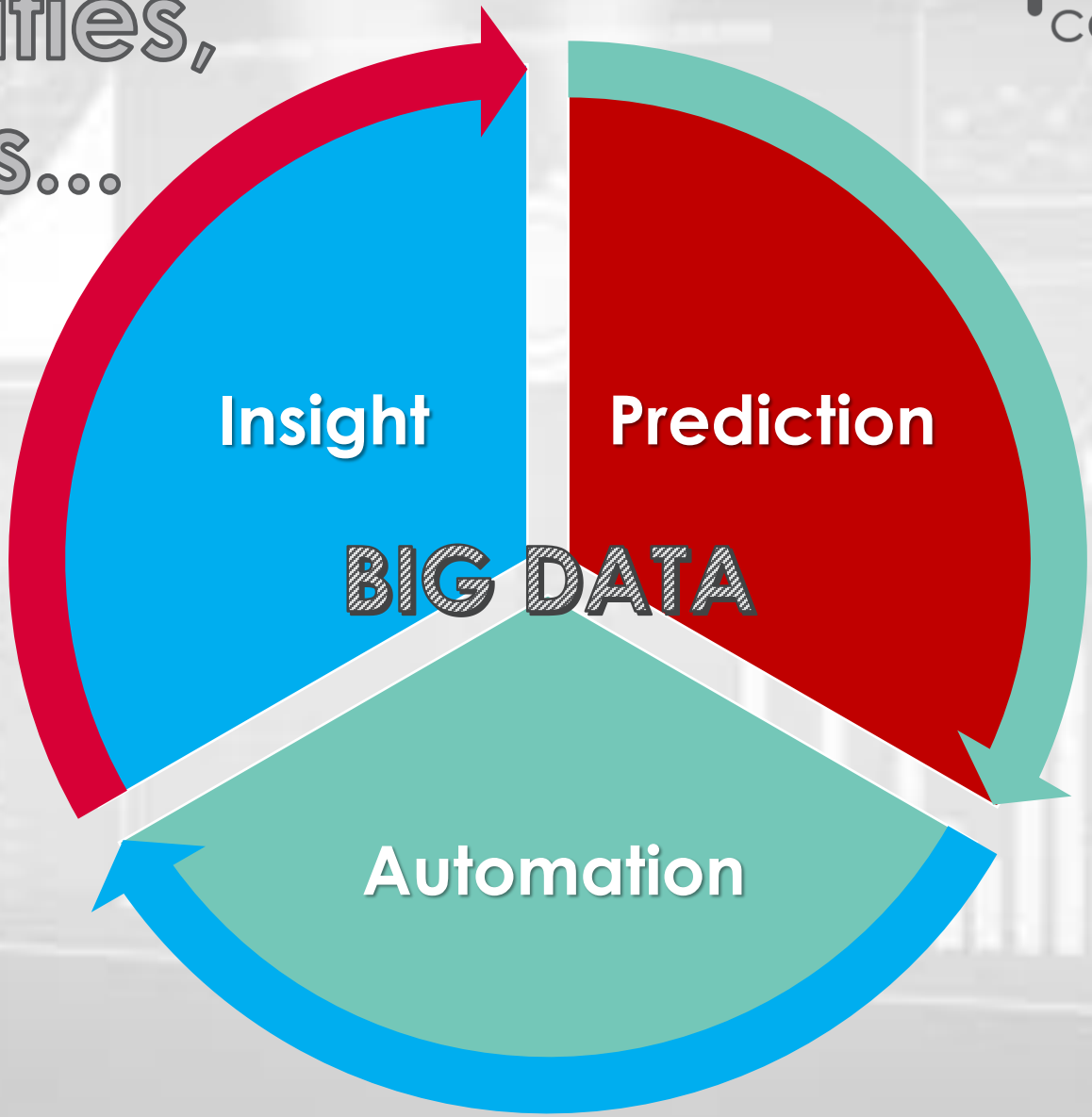
Source: ICAEW Thought Leadership Tech Faculty

The opportunities, and risks

- Three broad ways in which data is being used to improve business decisions
  - To gain insights (understanding)
  - To predict the future
  - To automate routine and non-routine decision-making

Source: ICAEW Thought Leadership Tech Faculty

Opportunities, and risks...



Opportunity 1: Insight

- Enhance understanding of operations, customers, risks, and markets, including:
  - New sources of data to gain new/enhanced information (more granular data about customers to understand their preferences, activities, location)
  - Exploiting the real-time nature of big data to improve services and operations (personalising responses and offers)
  - Applying analytics to gain new insights and interrogate entire data sets, e.g.:
    - Recognising new associations and patterns
    - Linking data from disparate sources to form a new/adjusted "big picture"
    - Identifying exceptions, unexpected behaviour, and outliers
- Real-world evidence, rather than assumptions, estimates, and guesses
- Improved business decisions about customers, suppliers, employees, strategy, and risk

Source: ICAEW Thought Leadership Tech Faculty

How have accountants benefitted from BDA so far?

- Data is at the heart of accounting
- Internal and external auditors
  - At forefront of the use of big data
  - Ability to analyse entire data sets
  - Changes in traditional approaches to audit based on sampling
  - Audit analytics help to identify outliers and exceptions, and focus on areas of greater risk
  - Wide range of analytical tools to visualise data, connect financial and non-financial data, and compare predictions to real world
- Accountants in business
  - Especially new sources of data and analytics tools
  - Improved forecasting (using real-time, non-financial data)
  - Linking financial and non-financial data can give better analysis of cost drivers
  - Detailed analysis of controls and operational processes
  - Identifying failures or pressure points which can be improved
  - Tools can be used in advisory roles, to help with business planning/operations

Source: ICAEW Thought Leadership Tech Faculty

Regulatory efficiency, and BDA (practical examples)

- UK govt's tax authority uses big data to combat tax fraud
- The Connect system draws on data from many sources, enabling sophisticated analyses to identify outliers and anomalies
  - Tax returns + data from internet + social media + land registry records + international tax authorities + banks
  - Crux = connecting data from disparate sources (NB!)
  - Development of artificial intelligence to do this for us
    - SA: "lifestyle audit" is pretty much still done manually
- SEC in USA: AI and BDA to identify insider trading

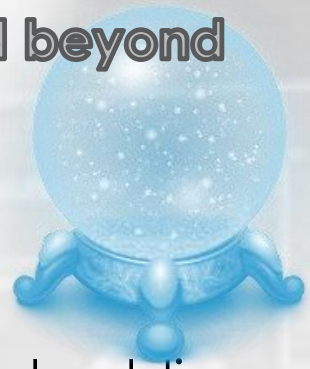
Source: ICAEW Thought Leadership Tech Faculty

Risks, despite gaining insight…

- Inaccurate, inconsistent, duplicated, out of date data
  - Amplified by big data, due to many sources, that could easily become outdated/unreliable (social media)
  - Data cleanses/complete disregard **vs** increased data volume that makes granular data quality less important?
  - Trade-off: volume of data **vs** accuracy of data
  - The context of use = NB (where quick response is needed, perhaps a lower level of quality could be accepted?)
- Sampling risks = redundant when big data is used, but broad conclusions must be justifiable
- Danger of correlations (beer vs nappies at midday ☺)
- Beware relying on averages, that could consist of highs and lows

Source: ICAEW Thought Leadership Tech Faculty

Opportunity 2: Prediction



- BDA enables prediction of future outcomes, embedding predictive models in business operations
- New sources of data
  - => New types of predictions
  - => Increased machine learning (AI)
  - => Greater reliance on prediction models
- Search engines can predict info most relevant to an individual
  - Other links to material, proximity to search term, other clicks on term, personalised (location, gender, age, previous searches), predictive suggested searching while you type
- Customer service predictive models personalise services based on predictions about individual customers (previous transactions, similar customers, qualities of products, when customer loyalty is likely to shift, responses to specific offers, etc.)
- Predictions about wear-and-tear of physical assets (scheduling of R&M, no disruptions, asset longevity etc.)
- Sports teams/coaches/commentators use predictive models to predict future performance of individuals and the team as whole

- **NB thought: Gaining insight versus prediction: how far can you reliably go?**

Source: ICAEW Thought Leadership Tech Faculty

Risks, despite predictions…

- Crux = reliance on associative correlations and identification of patterns, rather than a deep causal understanding
  - False positives and negatives
  - Conditions change, but underlying assumptions don't… could lead to inaccurate predictions
- Volumes of data could improve predictions, for example the prediction of the path of hurricanes (due to sufficient data), but not the intensity of hurricanes (as insufficient ocean data captured)
- Ethical concerns (refer next slide)

Source: ICAEW Thought Leadership Tech Faculty

Big data, and ETHICS…
(very relevant in practice)

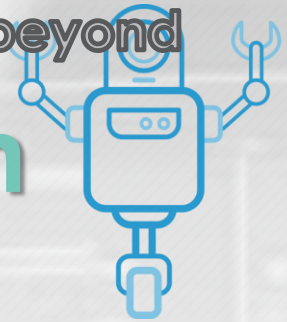
- Algorithmic-based decisions may be based on faulty models that give biased or unfair outcomes/predictions/recommendations
  - People classified/profiled based on similarities => used to predict their future behaviours

- Ethical concerns?
  - Not enough data to be statistically valid
  - Lack of feedback loop to pick up *errors* or *changes* in the environment
  - Transparent use of data, even sensitive data, that may otherwise not have been used…
  - Confusing correlation when compared to causation
  - Using cheap and easily-available data, rather than relevant high-quality data that may be hard/costly to capture and obtain

- Risks?
  - Discrimination against disadvantaged groups (e.g., criminal profiling)
  - Far-reaching impacts on behaviour, as people try to manipulate systems for their own advantage
    - The ratings of US universities: the selection of such ratings data drives universities to spend lots of money on their sports facilities to increase their rating, but doesn't give them an incentive to keep education fees low

- Source: *Weapons of Math Destruction,* by Cathy O'Neill

Opportunity 3: Automation



- The predictive capabilities are used to automate routine and non-routine decisions and tasks
- Phrase of the day: "RPA"
- Driverless cars
  - Data from sensors, mapping applications, satellites
  - Seen to improve road safety by removing human fatigue, carelessness and poor-quality or reckless driving
- Law and medicine
  - Healthcare companies: machine learning (AI) i.r.o. diagnoses
  - Computer can interrogate large data base much quicker than human being and consider all possible scenarios, based on presented symptoms
  - Use predictive model for most likely diagnosis
  - Legal: scan through large amounts of evidence for relevant info
- Accountancy
  - Bookkeeping and compliance work = automated
  - Machine learning (AI) starting to perform less routine tasks

Source: ICAEW Thought Leadership Tech Faculty

Risks, despite automation…

- Deep questions for professions, such as accountancy and law
  - How far can automation really go?
  - When are computers better, and when is human knowledge vital?
- How do we identify that predictions and/or automation have gone wrong? Who is responsible? Who will correct these errors?
- Less human interaction = increased risk of hacking and rogue computers
  - The importance of cybersecurity!

Source: ICAEW Thought Leadership Tech Faculty

cpd campus

How do we practically exploit[1] ([1]make full use of) big data? And what should we think of when doing so?

Big data exploitation

- Currently mainly by big companies, usually at the leading edge of data and technology, such as internet companies and major retailers…
- SMEs not benefiting from big data, as yet = opportunity!
- Let's look at some considerations of big data exploitation for…
  - Businesses working with data
  - Accountants
  - Policymakers

Source: ICAEW Thought Leadership Tech Faculty

Businesses working with data

- Where to start, especially with big data? Smaller businesses may not have much data, or struggle to identify what is useful
  - One approach is to start playing with data, and see what transpires, but experts recommend to start by defining useful business questions that need answers
  - Then possible sources of data can be identified
    - Sources of existing internal or external data
    - Some sources may not yet exist, but could be collected by tweaking systems or processes
    - Some sources may need longer-term planning, and a cost-benefit analysis
  - Data quality could be problematic
    - Data from different sources could initially be incompatible, require manual effort
    - Data clean up could be required, especially at first
    - Pervasive data standards across the business could be lacking at first
    - Significant time and effort to sort out the data, at first
  - As data experience grows, a shift is required to a data-conscious culture and building the right structures to maximise opportunities and manage risks
    - Shift from traditional decision-making cultures ("highest-paid person's opinion" to data-reliant)
    - Research indicates that entities that are data-driven, have higher output and higher productivity than those who don't
    - Tensions can arise from being data-driven (data can be overwhelming, result in paralysis)
    - Data overfocus could stifle innovation and risk-taking
  - Challenges around organisational structure, given the need to access many different skills
    - Building interdisciplinary teams = key element of success
    - Different approaches to this
      - Centres of excellence within the business to lead initiatives and share experience, working with other business areas as needed
      - Certain functional areas within business making use of real data, may lead organisational abilities (e.g., marketing, operations)
      - IT functions may take the lead

Source: ICAEW Thought Leadership Tech Faculty

Businesses working with data

- Skills required to exploit big data and analytics
  - Statistical skills to build algorithms and understand the robustness of models
  - Data and technology skills to extract and manipulate data
  - Domain knowledge to ask the right questions and gain insights from the analysis
- Some skills can be bought from third parties (data analytics service providers), rather than attempting to build high levels of skills internally
- Cloud model (access technical resources without investing in substantial hardware)
- Cross-organisational ownership and sponsorship of data projects could be challenging (focus is usually on own data and projects)
- Controls and governance in place to ensure that data models are used appropriately, and to create sustainable value?
  - Many governance issues should be considered
  - Rule-based (simple) versus machine-based AI (algorithm constantly evolves, who understands working and assumptions?)
- Data and algorithms contribute to corporate value – investors may want assurance around their long-term sustainability…

Source: ICAEW Thought Leadership Tech Faculty

Accountants

- *Seen to be in "early stages" of use of big data and advanced analytics, significant opportunities still to be realised…*
- Opportunity for accountants to transform their role into wider "guardianship" of data across organisation
- Discipline, structure and ethical approach places accounting profession well to assist with effective use of big data and analytics
- Accountants' natural prudence and skepticism ensure robust and appropriate use of big data
- Marketing, IT and operations specialists could displace the accountant in a more dominant role around big data
  - The same applies to data scientists that could marginalise the accountant in decision-making in case of the failure of the accounting profession to keep up with developments around data and related technology
- Risk that financial reports may become less important to investors and other stakeholders as the impact of big data grows
  - Predictive models for profitability; wide range of "alternative sources of data"
  - Audited info always important, but profession could become marginalised to decision-making if oblivious
- Accountants need to be upskilled in statistical analysis, sufficient to be 'good buyer' and ask good questions of suppliers and other parts of the business
- Accounting qualifications are evolving to include a focus on data and related data analytics
- A pervasive mind shift is also required – playing with data to "see what comes out"; using data to interpret reports as value-added service; testing ideas with real-life data; etc.

Source: ICAEW Thought Leadership Tech Faculty

Accountants

- Which skills do **accountants** have to develop to effectively work with data scientists?
  - Understanding the opportunities and limitations of what can be achieved through data and how data science can add value
  - Understanding and defining the business problems data can help to solve
  - Interpreting the outputs produced by data analytics. This includes understanding data provenance, modelling assumptions, inherent biases in the analysis and, perhaps most importantly, what decisions can justifiably be made based on the analysis
  - Presenting and communicating the results to the business, including the use of visualisation
  - Awareness of the data landscape, different data types, what data might be useful and where and how it can be obtained
- Research into big data and the profession has also emphasised the importance of business and commercial skills, and the opportunity for accountants to play a bridging role between data science and other business functions
- *As a result, while data and statistical skills are important, they are only part of the bigger picture*

Source: ICAEW Thought Leadership Tech Faculty

Policymakers

- *Governments are some of the biggest users of IT systems and generate great volumes of data about citizens and services*
- Opportunities to improve/personalise services, target resources/interventions, and inform citizens
- Evidence-based decision-making by enabling deeper analysis of the impact of policies
- Enhanced democracy and transparency through the release and use of open data
- Creation of skills across the economy to enable widespread use of big data; concerns about the shortage of skills of data scientists in future
- Not just specialists but also pervasive ability among workers to use quality data to make proper business decisions, understand prediction models' working, risks, inappropriate reliance, etc.
- Creativity and imagination are important to ask the right questions in appropriate areas – this should be fostered, and experimentation encouraged
- An appropriate regulatory framework should be carefully considered!
  - Patterning and profiling of individuals = deep social implications
  - When is it acceptable to judge people based on data about past behaviour, personal characteristics, and similarities to others?
  - Strongly linked to debates about privacy: deep insights into individuals
  - Benefit is currently cheap/free personalised services
  - How long will consumer be content with this trade-off as analyses go deeper into our activities/personal lives?

Source: ICAEW Thought Leadership Tech Faculty

cpd campus


Thank you
for attending
sait
South African
Institute of
Taxation
"To adapt to our complex world of weaponized information,
maybe schools should teach data as we do languages."
— Roger Spitz