George Hastings

As the minds of machines evolve at a disturbingly breakneck speed into uncanny manifestations of the musings of science fiction, it grows ever inevitable that data analytics become irreparably entangled in the socioeconomic dynamics of modern society. In fact, Max Tegmark said in *Life 3.0* that "The saddest aspect of life right now is that science gathers knowledge faster than society gathers wisdom" and with this generational escalation in artificial intelligence large corporations are unavoidably beginning to utilise this software for their own (often nefarious) purposes. So to what extent are data analytics already being used to influence the minds of commonplace consumers and how far can they go in fundamentally altering the dynamic of both commerce and culture? At an even more fundamental level, we will discuss the fundamental mechanisms behind data analytics as big data's insurmountable value is gradually reflected in infrastructure and politics.

Right now, around the world, people are being influenced through data analytics. But in ten years such primitive techniques may be discarded as obsolete, with artificial intelligence pioneering this shift towards far more impersonal management and distribution of data. As the 2030 deadline for vehicular electrification draws inexorably closer, as will the impetus on self-driving car ownership escalate, propagating machine learning to the forefront of industry and in turn forcing a plethora of firms into the broad world of machine learning. The result? A workforce capable of using their newfound coding skills not only for ingenuity in the mechanical sectors but also in other areas, predominantly marketing. This symbiosis between machine learning and data analytics will mean influence broadens as seemingly sentient algorithms become capable of viewing inexplicably large data sets and condensing such monoliths into applicable representations that can be used to influence. Already, *machine learning in the form of algorithmic advertising has generated billions of dollars in profits for private companies, but the technology has accomplished far less in the public sector* (OHIO's Online Master of Business Analytics), and this is revealing of the foundations of corporate artifice; our data is simply not safe and will be used against us.

George Hastings

When talking about how data is used to influence us, we must consider whether such analytics are influencing consumers for better or worse; will advertisements based on machine learning simply be a tool which show us what we want to see? Or will it be used to pervasively implement a late-state capitalistic society as consumers are unwillingly coerced into buying products or services from the tech giants that frequent silicon valley? Organisations such as International Aid Transparency Initiative promote transparency in the use of data resources but one wonders how much impact a more holistic approach to data will actually have; the popular Netflix film "The Great Hack" gave a glimpse of just how insecure data is, describing how up to 87 million Facebook profiles were leaked by Facebook to the data analytics firm Cambridge Analytica for the purpose of tampering with the 2016 Cruz-Trump campaigns. The astounding story of Mark Zuckerberg's willingness to selfishly disclose personal information of millions can be used not only to show the risks of how data could be mistreated in the future, but also highlights just how much your data is being used to influence you already.

The power of machine learning in respect to data analysis is leveraged and limited by hardware; as Gordan Moore's 1965 proposal that transistor counts would double each year shows signs of slowing down after 50 years of accelerated semiconductor integration, we could turn to the continually increasing power of co-processors like graphics cards to process massive amounts of data in a parallelised fashion, especially when using task-optimised hardware; leveraging the bountiful resources of modern GPUs must not be ruled out. Yet unless we fundamentally redesign processor architecture the demands of machine learning will likely still surpass the capabilities of modern hardware. This would call for serious software optimisation as well as customized mathematical processing engines. But why invest extortionate amounts of money into data analytics when our brains have an average of 86 billion neurons each? Andrew Ng, who is not only a leader in the field but also an accomplished teacher (his Stanford course, CS229, is testament to this) has recently propagated this notion which has taken the machine learning community by storm: one can minimise the various biases (for example confirmation bias, where outcomes confirm existing assumptions and prejudices, or stereotype bias wherein unrepresentative training data causes false correlations in models) which could sabotage machine learning models by

simply foregoing the several stages of a complex neural network. To take a more general example, if one wanted a computer to recognise a handwritten letter then multiple layers of interwoven neural networks would be required to provide an answer, whilst a human could execute this computationally complex algorithm in an instant. How does this apply to data? Some correlations in datasets are highly interpretative, and thus the role of the human in data analytics still shows great promise. One would imagine that, in particular, humans will remain indispensable in deciding how to enact upon data analytics since, whilst intelligence of technology is being continually expedited, it lacks the subtlety that a human hand can bring to the table. In essence, the boom in technological development along with the bounds forward in software development will mean that from a data analytics perspective influencing consumers will become ever easier. Let us just hope that we do not beckon our own demise by handing over apodictic control of our data to the AIs we once nurtured into existence.

But when presented with flurries of correlations and inferences our silicon sidekicks have conveniently made for us, how best do we present this raw data with beneficial strategies which, from the perspective of an enterprise, will convert to revenue? Various strategies can be employed other than the obvious investment in the aforementioned advanced analytics competencies (data science and machine learning), for effective software systems truly are only a building block upon which great data analytics can be formed, accelerating and automating the process. Drawing valuable insights from data analysis requires the supervision of a data analyst, in particular to tackle qualitative problems with subjective variables like drawing inferences from twitter responses or Instagram stories which, however sophisticated our algorithms, contain a level of nuance suitable only for a human professional. It is better to think of machine learning and data science resources not as replacements with which we can one day automate data analysis, but rather a tool enabling us to more proficiently fulfil the "three V's" proposed by Doug Laney (industry analyst): Volume, Velocity and Variety. In particular, computers are extremely beneficial for increasing velocity whilst they are somewhat hindered when it comes to variety. Nevertheless, they

still do offer capability when it comes to both quantitative (their forte) and qualitative (more subjective – their weak point) data since fundamentally they can handle any task as long as it is sufficiently abstracted into graphical or numerical form.

Therefore computers, can fundamentally only work with dependent and independent variables which it can plot on the y and x axes respectively. These number-crunching experts are better used automating certain repetitive processes like the regular refreshing of datasets or the removal of anomalous data so that the job of the humans is simple: to influence as many people as they can, for better or for worse. It is also fundamental to consider how corporations can skew the representation of data in their favour; employee wages in a company can easily be misrepresented if the salary of the CEO and other executives is included when calculating the mean. Equally, if the magnitude of the dataset is limited then, and this is particularly prevalent in the medical sector, data can be presented in multitudinous ways which draw false conclusions from the data.

Data analytics is a process which is indispensable within the marketing and advertising sectors, and the automation and rapidity provided by software tools is undeniably important. However, it soon becomes clear that the level of investment required to implement a machine learning ecosystem to repeatably analyse data autonomously is often outweighed by the simplicity of hiring "good old" data analysts. Certainly, this role is by no means obsolete, with demand rising 231% over the past five years as corporations realise that the best method of using data analytics to influence people is through a symbiosis of man and machine rather than a surfeit of one or the other; the rare human gift of emotion, which it is doubtful any other animal on earth possesses, can be fused seamlessly with our big Boolean brother, the computer, to not only recognise trends but also to present this data in a manipulative yet substantiated fashion. I, for one, am far from frugal when it comes to handing over personal information as visions of a society whose very foundations revolve around data and its distribution appear day on day to be closer reflections of reality.