What is the Impact of Big Data on Business, Workforce & Society?

Ben Mullally

Audience: WIRED Magazine Editor

Note for WIRED Editor

I think a lot of people are hearing the words 'Big Data' talked about but still do not know what it is. By producing this report, I hope to explain it to them and give them a wide variety of case-studies of applications, hoping to touch upon something that affects their lives. This is so they can understand how the world is evolving with data and prepare them for the future, whether it is their jobs that are affected, the city they may live in or many other factors in life. I think a major concern is that ordinary people do not understand how much of their private information is out there being used for analytics. It is important to wake people up to this scary truth and remind them of the dark sides to this technology too.

What is Big Data?

Big Data is quite simply extremely large amounts of data that have been collected. Just about every action you complete on the internet is recorded such as sending emails, website clicks, searches, social media likes, etc. It can be analysed to find patterns, trends, and relationships, especially related to the way humans act and interact with one another. The term was first coined in the 1990s and ever since its relevance has continued to grow. From 2010 to 2020, data usage increased 5000% from 1.2 trillion gigabytes to almost 60 trillion gigabytes.[1] According to Statista, this number will rise to 180 zettabytes by 2025. This is gold waiting to be mined.

Traditional software cannot process the vast quantity of new information, so techniques have been developed and implemented over the years by researchers in the field. The main tool, and likely to be the future of all technology, is machine learning. This is an area of artificial intelligence whereby a system can automate the modelling of a data collection and its reliance on large amounts of data makes it a primary use in data analytics. This technique can be applied through three different types of method to predict future outcomes:

- Supervised learning: models learn from training with labelled data using cross-validation (repeated split-sampling over limited dataset) to group data (clustering).
- Unsupervised learning: models group unlabelled data (clustering) by common associations.
- Semi-supervised data: models originally established using labelled data and uses correct predictions on new data to refresh the model to increase its accuracy and improve adaptability.[2]

Deep-learning models use large amounts of data of multiple dimensions with a hierarchy of supervised and non-supervised techniques making it perfect for analysis of Big Data. Descriptive statistical methods are used to give summaries of data and then inferential statistics can discover the behaviours of a population through samples and statistical techniques such as regression analysis, whereby relations between different forms of data are established. The combination of machine learning, data mining and statistical methods are the fundamentals for Big Data analysis and ultimately lead to data-driven solutions for companies, governments,

hospitals etc. This report will focus on the impacts of Big Data by separating them into three major categories: Business, Workforce and Society. These will then be comprised of further subtopics that give specific case studies and examples of the applications of Big Data in the real world.

Applications of Big Data

Business

Most organisations are now starting to use Big Data analytics to focus on their marketing and cost-cutting strategies to advance global efficiency, giving them a competitive advantage over the rest of the field. Policies have even been created by governments to promote its use due to the obvious economic and societal impacts. For example, on 9 September 2020 the UK government launched its 'National Data Strategy' which has plans to integrate data into our lives to improve quality of life.[3] From New Vantage's 2022 survey, 97.0% of the 94 Fortune 1000 companies questioned were investing in Data initiatives and 91.0% were investing in AI.[4] The top companies in this survey were mainly from a US finance/insurance or healthcare background. As they have extreme wealth and power, they are choosing to invest early into data.

Understanding customers

Using the information of user's buying habits, organisations can tailor their product to the market whether it is through small improvements such as software updates or completely new ideas for products to sell. This makes the old lengthy ways of taking large customer surveys obsolete and as a result less resources (time, money, etc.) are wasted. For example, Amazon whole foods uses the data acquired from customers buying habits and the exchanges between the suppliers and grocers to find insights on what changes need to be made.[5] Furthermore, Netflix is an easy example of how a company uses your data (search history and watch data) to target adverts for new shows towards you whilst you still use the app. Big Data can also be used to boost customer retention and acquisition, risk management and supply chain management; all of which give a company a competitive advantage. Companies no longer rely on customer feedback because they can use current data analysis techniques to predict what is desired from a product.

Reducing cost and improving efficiency

Big Data analysis makes it possible to predict and forecast the patterns of sales in a company's market allowing for them to plan the production of their product. This includes how much they need to make and when they should make it for. Profit and customer fulfilment can be kept as high as possible without unnecessary costs. This massively reduces risk and improves efficiency; if everyone in the company knows they are maximising their time, it is likely to improve employee resourcefulness and happiness. PepsiCo is an exemplary company that prides itself on using Big Data to manage their supply chain to keep it as efficient as possible.[6] It obviously keeps investors happy too as the company is more stable with statistics backing up its goals, supporting the Theory of Constraints. This is the idea that finding the restricting factors of a business and then focusing on improving them methodically until they no longer limit can help you reach your overall goals. Production and sales are the obvious departments that can be focused on however more-in-depth analysis of employee workflows, analysing social media interactions, targeted advertisements etc. all cut costs and increase the efficiency of a company.

Problems

Despite the clear benefits of big data, even Fortune 1000 organizations continue to struggle to become data-driven, with only 26.5% reporting have achieved this goal, and only 19.3% reporting having established a data culture as of 2022.[4] The field is still in its infancy and many challenges have been identified.

Firstly, 80-90% of the data we generate today is unstructured and 95% of businesses allude to the management of this as a problem, and so heavy investment is needed from companies to correctly store it all.[1] There are several different types of state of the art storage solutions including Distributed File Systems, NoSQL databases, NewSQL databases and Big Data Querying Platforms but Cloud Storage and 'data lakes' are the predominant solutions.[7] These are all still extremely expensive and time-consuming investments, especially when all this data must be kept secure afterwards too. Secondly, 91.9% of top businesses acknowledge that cultural challenges are the biggest obstacle to becoming data driven.[4] This ranges from changing business models, to outdated administrative structures, and companies not risking change in the way they are run. Finally, companies need to invest in the technology and train the people who will run this side of the business. This will be touched upon later.

Every company now knows that data is the future, but it is proving a fight to adapt to the new technology. Almost every industry can and should be using Big Data to advance, but it is human limitations that are currently holding them back, not the technology.

Workforce

New jobs

As the demand for data increases, businesses need qualified people to analyse this data. Big Data is useless to a company if they cannot understand or analyse it and so this has spawned brand new occupations in the last decade.

The huge amounts of information that companies are now buying and collecting need to be stored in expansive databases, generally in outsourced cloud-based or physical storage warehouses. This has created new departments in companies, like Amazon's S3 storage service, and even whole new companies like Hadoop that specialise in providing data storage and parallel computing of datasets for other corporations. These databases need to be managed by Database Administrators who are responsible for controlling access, backups, and recovery. Data Architects work with every aspect of the department to ensure the databases can be centrally implemented with high security and providing system reports to keep everything up to date and working smoothly. The Big Data ecosystems and the efficiency of the databases need to be constantly built, updated and tested by Data Engineers to allow for Data Scientists to run their algorithms on secure and optimized systems. The scientists are very qualified in data analysis and processing to make predictive models on unstructured data to identify trends and patterns that could be of use to the business. Data and Business Analysts also handle data focusing on visualisation and optimization to give insights on the company and what it can do to improve. All these jobs are in demand and generously paid; a booming new job market has very quickly been created for young professionals.[8]

Threat of Automation in the workplace

The workplace has been positively impacted by the use of data analytics to give better hiring decisions, improve employee efficiency and happiness, and help HR find problems in offices via AI pattern finding. It is extremely important to note that this new market is still very reliant on humans to run it. The media loves to exaggerate the autonomous takeover of people's jobs and

although machines are undoubtably replacing people in places such as factories and warehouses, there are new professions being created that are in demand. For example, Data Science related jobs are predicted to rise 22% between 2020 and 2030, according to the US Bureau of Labour Statistics compared to the average US occupation growth of 8%.[9] Every statistic point to the number of Big Data related jobs to continue to skyrocket. Miguel Alcaine, head of the International Telecommunication Union's area office, Tegucigalpa, Honduras, wrote "If some high-placed people believe this type of technology can predict the unpredictable, there will be cases where this technology will be overextended and misused. Human judgment cannot be replaced by technology, the former being the responsible for decisions".[10]

Society

Healthcare

Big Data has had some huge impacts on the healthcare industry and the people using it. Firstly, data can now show patients how to take an active role in their health through their diet, exercise, and medication. The future could show smart home items such as toothbrushes, toilets and scales giving instant feedback on the state of your health. Secondly, Big Data tools can collect millions of records on 'historical patient data, real-time data from monitors, clinical factors, lifestyle choices and social determinants' to enable evidence-based treatment specific to the patient and suggest the best solutions to professionals in seconds. Furthermore, real-time data has been used to recognize crisis before they even happen. Lastly, healthcare has become more cost-effective through utilizing Big Data on "patient-outcome compensation" and removing "fraud, waste, and abuse" in a system. [11]

In 2020, health-care expenses represented 19.7% of the US GDP and it is expected by 2028, the spending will reach one fifth of the nation's GDP (about $6.2 trillion).[12] This rise is happening in every other country also, although not as extreme, however Big Data is already being harnessed to battle medical costs. For example, The Clinton Health Access Initiative (CHAI) has used analytics to forecast the demand for certain disease medications such as HIV/AIDS, malaria, and tuberculosis which led to talks for decreased prices and wider availability of medical care in the third-world countries.[12] They also share their data with the UN and WHO to help identify the best way to use their funds. Furthermore, tools are already being used in Queensland hospitals in Australia to accurately predict the number of patients and their injuries for years in advance. Using historical admission and discharge data to do so, the tool allows departments to prepare for beds, staffing, surgeries, and medications. It involves spotting spikes in patient activity through Box-Jenkins Autoregressive Moving Average Analysis, a type of time-series statistical analysis, and then forecasting future activities. This has cut waiting times and cancellations as much as 20%. As of 2017, it had already saved the hospitals $2.5 million each year and countless lives but the value to the state is likely around $77.5 million.[13]

Government

Governments are already using Big Data to enhance several aspects of society. It can be leveraged to project weather forecasts, predict natural disasters, city/public planning, traffic organisation, personalised education, fraud detection, automatic translation, smart homes, etc. Here we will focus on a few interesting examples already in use.

Many governments are heavily investing in smart-city data solutions to handle large city growth in population and living costs. Barcelona has implemented public Wi-Fi mesh networks with new heating, cooling and energy systems that is part of a 'smart metering program', based on the collection and analysis of public information, to make the city more sustainable and provide

high-speed internet connectivity between citizens.[14] Moreover, TfL and Siemens partnered for a 10-year contract in 2018 to deliver a massive upgrade to London's traffic management systems that could have £1 billion of benefits by reducing delays. Glynn Barton, Director of Network Management, TfL, said "It will use new data sources to better manage our road network, tackle congestion, reduce delay for people choosing healthier travel options and improve air quality".[15]

Additionally, Estonia has created a data exchange network called 'X-road' which approves database requests and automatically shares material between government agencies without manual requests needed from officials or citizens. It uses strict rules on protocols to ensure data availability, confidentiality, and integrity. This apparently saves the country's administration "more than 240 hours of work every three minutes" and has been so effective that Finland has implemented the system too. In 2018 the two countries even connected their networks to further develop the X-road core[16]

Security & Privacy

Urban security has been improved by helping enforcement agencies to predict crime before it happens. By using data on crimes in certain areas and particular buildings, 'Risk-Terrain Modelling' can be used map the locations of an area with high crime rates and can flag to local enforcement agencies where different crimes are being committed. This can let police prepare in advance for dangerous crimes by having the resources and staff available to quickly respond in high-crime areas. The system is simple and is already being used by over 45 countries where many cities saw reductions in crime by over 30%.[17]

On the other hand, Big Data does not work without the collection of our information; the more it has the more useful it can be for our lives, but is it worth losing our privacy for? People simply do not know how much of their personal information is out there to be used against them. A current threat is from 'Algorithmic profiling', a method used by data brokers to target parts of society by their level of vulnerability and then illegally sell their information (e.g. rape victim files, patients with genetic diseases, and addicts). Furthermore, credit card companies have been caught using people's personal information to set rates and limits without the them even realising.

Algorithms that use Big Data can also discriminate unintentionally, as the factors that make up a human such as skin colour, sexuality or gender are just parameters in a machine processing a result. For example, in 2017, an extensive study by ProPublica and Consumer Reports gave evidence that "minority areas have to pay higher car insurance than white areas with the same risk".[18] Additionally, some studies show targeted ads are showing more higher-paying jobs to men than women in a climate where there is already tension over the gender pay gap.[19] Risk-Terrain Modelling was praised earlier due it's obvious benefits for police departments however this can lead to the over-monitoring of poor communities. More police are drawn into an area which can firstly cause more tension and lead to more crimes recorded in these areas "creating a pernicious feedback loop".[18] These examples raise the question of who oversees these algorithms and what could happen if somebody with bad intentions was to take control and discriminate even more?

What does the future hold?

It is clear that Big Data is at the forefront of business, workforce, and society but each have their own problems...

In terms of business, the majority of companies are now investing in the technology for analysis, but it is still a slow overall progression. The larger corporations are targeting data as much as they can to get ahead of their competitors however small businesses will take decades before they can catch up due to the huge shift in culture that is required. Every company will eventually have a department for data analysis, with the more powerful ones using their resources to focus on Big Data. Furthermore, Big Data departments within these companies are going to be so valuable to them that it is likely that they become more important and have more investment than their primary divisions, such as sales and manufacturing. This is similar historically to how the sales departments of certain businesses became larger and more valuable than the actual manufacturing and quality of the product. It is just the natural evolution of business and those that don't quickly adapt to Big Data will simply be left behind.

With Workforce, massive new job markets are being created and it is evident that they are going to be sustainable, with salaries being generous too due to the little experience there currently is. For now, automation and data analysis of the workplace is generally a very positive thing, especially for the western world, and it will continuously advance the efficiency of offices, factories, hospitals etc.

In conclusion, it is apparent that Big Data has had the biggest impact on Society so far. There is an overwhelming number of positive influences it has already had, and, at this moment, they far outweigh any side-effects. Quality of life is constantly being improved with new analytical research and development making work-life easier, helping governments develop their countries and even aiding hospitals to save more lives. It is however slightly worrying how effective Big Data is when used properly, and so it raises concerns to whether its use is of moral good with only the most powerful currently in control of it. Long-time technology analyst Oscar Gandy summed this up perfectly: "If 'Big Data' could be used primarily for social benefit, rather than the pursuit of profit… then I could 'sign on' to the data driven future".[10] It seems that all massively lucrative corporations nowadays have had huge scandals leaked, creating trust issues with the public; some believe they would not be as successful if they had followed the rules. Unfortunately, they are also the companies capitalising quickest on the Big Data explosion.

References

1. Kulkarni, RK. 2019, 'Big Data Goes Big', Forbes, February. Available at: https://www.forbes.com/sites/rkulkarni/2019/02/07/big-data-goes-big/?sh=7da7397420d7
2. Juan-Pedro Cabrera-Sánchez, Ángel F. Villarejo-Ramos, Acceptance and use of big data techniques in services companies, Journal of Retailing and Consumer Services, Volume 52, 2020, 101888, ISSN 0969-6989, (https://www.sciencedirect.com/science/article/pii/S0969698919302589)
3. UK Department for Digital, Culture, Media & Sport (2020) National Data Strategy. Available at: https://www.librarydevelopment.group.shef.ac.uk/referencing/harvard.html#EFG
4. NewVantage Partners (2022) The Quest to Achieve Data-Driven Leadership: A Progress Report on the State of Corporate Data Initiative. Available at: https://www.newvantage.com/_files/ugd/e5361a_ad5a8b3da8254a71807d2dccdb0844be.pdf
5. Kopanakis, J. (2018) '5 Real World Examples of How Brands are Using Big Data Analytics', Mentionlytics, 14 June. Available at:

https://www.mentionlytics.com/blog/5-real-world-examples-of-how-brands-are-using-big-data-analytics/
6. Tableau Conference (2018) How PepsiCo's Big Data Strategy is Disrupting CPG Retail Analytics. Available at: https://www.tableau.com/learn/webinars/how-pepsicos-big-data-strategy-disrupting-cpg-retail-analytics
7. Strohbach, M., Daubert, J., Ravkin, H. and Lischka, M., 2016. Big data storage. In New horizons for a data-driven economy (pp. 119-141). Springer, Cham.
8. Chatterjee, MC. 2022, 'Top 9 Job Roles in the World of Data Science for 2022' , Great Learning, January. Available at: https://www.mygreatlearning.com/blog/different-data-science-jobs-roles-industry/
9. US Department of Labor (2022) Occupational Outlook Research Scientists: Computer and Information Research Scientists. Available at: https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm#tab-1
10. Janna Quitney Anderson, Lee Rainie, Pew Research Center's Internet & American Life Project (2012) The Future of Big Data. Page 2. Available at: https://www.pewresearch.org/internet/2012/07/20/main-findings-influence-of-big-data-in-2020/
11. Kumar, J. (2016) 'Impact of Big Data Analytics on Healthcare and Society', Journal of Biometrics & Biostatistics, 7:3. DOI: 10.4172/2155-6180.1000300
12. Yang, J. (2022) U.S. national health expenditure as percent of GDP from 1960 to 2020. Statista. Available at: https://www.statista.com/statistics/184968/us-health-expenditure-as-percent-of-gdp-since-1960/#:~:text=In%202020%2C%20U.S.%20national%20health,higher%20than%20other%20developed%20countries.
13. Apolitical (2017) Queensland hospitals can predict admissions ten years in advance. Available at: https://apolitical.co/solution-articles/en/queensland-hospitals-save-millions-forecasting-patient-admissions (Accessed: 8 February 2022)
14. I. Yaqoob, I. A. T. Hashem, Y. Mehmood, A. Gani, S. Mokhtar and S. Guizani, "Enabling Communication Technologies for Smart Cities," in IEEE Communications Magazine, vol. 55, no. 1, pp. 112-120, January 2017, doi: 10.1109/MCOM.2017.1600232CM.
15. UK Department of Transport (2018). Delivering the next generation of urban traffic management. Available at: https://tfl.gov.uk/info-for/media/press-releases/2018/june/delivering-the-next-generation-of-urban-traffic-management
16. Republic of Estonia Information System Authority (2021). Data Exchange Layer X-tee. Available at: https://www.ria.ee/en/state-information-system/x-tee.html
17. Kennedy, L.W., Caplan, J.M. and Piza, E.L., 2015. A multi-jurisdictional test of risk terrain modeling and place-based evaluation of environmental risk-based patrol deployment strategies. Newark, NJ: Rutgers Center on Public Security.
18. J. Angwin, J. Larson, L. Kirchner and S. Mattu, ProPublica (2017) Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk. Available at: https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk
19. Anja Lambrecht, Catherine Tucker. (2019) 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads', Management Science 2019 65:7, 2966-2981. doi: https://doi.org/10.1287/mnsc.2018.3093