What is Big Data Analytics and Why is it Important?

There is no definitive definition for big data analytics, it is more of a concept. The concept of big data analytics has been around for years. Most modern and successful companies know how to use data that enters their business can be used to their advantage. These organizations can obtain significant value once you apply analytics to these chunks of data. This idea of using data to discover trends and find insights was even utilized many years ago. They used countless spreadsheets instead of the modern computer to discover and take advantage of these trends. The modern way of reaping the benefits of big data is more efficient than ever. “The ability to work faster – and stay agile – gives organizations a competitive edge they didn’t have before” - SAS Insights.

Big data analytics is important for three main reasons: The introduction of new products and services, cost reduction and, as mentioned before, faster and improved decision making.

Companies now have the necessary tools to monitor customer habits and needs using big data analytics. These organizations can find out what satisfies the customer so they can give them what they want. This is clearly a huge benefit to the customer as well as the business. Research from Tom Davenport “points out that with big data analytics, more companies are creating new products to meet customers’ needs”. Tom Davenport is the IIA Director of *Big Data in Big Companies*, which is his report on big data analytics. Davenport interviewed more than 50 companies to help better understand how they are using big data.

Cost reduction is possible with big data analytics as efficiency is increased. Typical processes needed for production such as testing, and quality assurance can greatly affect industries like nanotechnologies and biopharmaceuticals. By using big data analytics, companies can make calculated and improved decisions as they discover the many issues they may face with their product. It also goes without saying that this saves time and energy.

The speed at which an important decision can be made using big data analytics is a huge reason why large companies can be up to date and successful in their respective industries. They can make game changing decisions to not only increase production but improve the quality of production. “This gives them a competitive edge and provides a more agile framework for decision making and risk handling” - New Generation Application.

(Page by: Adam Parker; ID - 19258356)

Sources:

-SAS - <https://www.sas.com/en_ie/insights/analytics/big-data-analytics.html>

-*Big Data in Big Companies –* Tom Davenport

-New Generation Application - <https://www.newgenapps.com/blog/what-is-big-data-analytics-benefits-challenges/>

How Companies Collect Big Data:

“Collecting customer data has become a major priority for businesses” - IT Chronicles. It is possible for companies to establish and take advantage of trends using data and the capturing and use of this data has become increasingly efficient as the technology for it has developed and improved.

Whenever you enter or use a website online, you are usually asked to consent to something called cookies. Well, this is one way in which companies collect your data. Cookies track activity on their respective website. Think of your cookie as an ID card. Each time you enter a website, you add to your “Cookie ID card”. Companies can use the information from these cookies to learn what sites customers most frequently visit and analyse this for their benefit. Businesses can also use this method to make their websites more practical and quicker to purchase products from. They do this by recording your personal information, so you don’t have to enter it multiple times each time you buy. Personalized ads can also be produced using the data in these cookies, e.g. if you are consistently searching for a new pair of running shoes, your “cookie” will record this data by analysing the sites you have visited and display ads for Nike, Adidas, Asics, etc.

Most companies will directly ask their customers for their data. This happens whenever you’re asked for your name, email address, home address or other information about yourself. This usually takes place after you purchase a product or subscribe to a service when you need to fill a form out. The company uses this information to send you ads and can use your name to personalize them. As mentioned before, companies can also store this information in cookies to use in the future.

Big data analytics also comes in the form of customer reviews. Many companies will insist on customers writing a review of the product or service the company had provided. Companies can use this information to improve the quality of their product or service.

One of the easiest ways in which companies can acquire your data is by simply searching through their own records. A company’s customer interactions and transactions are filled with useful data which can be analysed to be used in to make business decisions. This can also be done using social media. “If customers use, say, their Facebook account to log into a third-party application that a company may provide” - IT Chronicles. Social media is open to anyone online so any data you post, or share can be accessed by companies. This method is simple and effective for large companies.

Big data analytics has become so popular amongst many industries that there have been data companies set up to find, purchase and sell on customer data to other companies. An example of one of these companies is Acxiom. Acxiom is a company with one purpose: “Collecting, analysing and selling customer and business data for targeted advertising campaigns”. Another company who operates similarly is Oracle. Both Acxiom and Oracle acquire their desired data from smaller data companies or even government censuses. The following is quoted from Oracle’s ‘AddThis Audiences: A Buyer’s Guide’: “*Collection Methodology: AddThis aggregates the online actions taken by unique visitors on every page across their vast publisher network. Our proprietary blend of observed (e.g., arriving on site from search, content on page, reading an article, scrolling, watching a video) and declared (e.g., click, like, share, follow, print) data yields a holistic picture of the true web habits, interests, and preferences users reveal when browsing online. We deliver unmatched audience insight and unparalleled audience segment quality at scale.”*

Acxiom has a similar ‘Buyer’s Guide’ which is full of information such as data types, and other data companies.

(Page by Adam Parker; ID - 19258356)

Sources:

-IT Chronicles - <https://itchronicles.com/big-data/how-do-big-companies-collect-customer-data/>

-Indiana University - <https://kb.iu.edu/d/agwm>

-Acxiom - <https://www.acxiom.com/>

-Oracle - <https://www.oracle.com/index.html>

The Future for Big Data:

Big data is the mass amount of data sets that are collected and gathered by companies, technology and a large amount of other sources .Big data is the future of modern businesses and marketing and will play a huge role in companies marketing plans using this data to understand upcoming trends and make a profit. A large quantity of this data is unused and unprocessed, so this area must be developed in the future.

“The big data analytics market is set to reach $103 billion by 2023” (Petrov, 2019). This is an estimation for the future value of big data, this value will only grow as big data becomes more important for companies marketing and research. With the creation of new technology and the growth and popularity of them, like smart watches or smart home devices, more data will be generated and data will be used effectively to understand trends and consumers wants, meaning this data will become more valuable as technology increases to grow.

Due to the growth of technology and the internet big data is becoming more and more relevant and important. Big data has countless ways it can be used, with technology developing more effective and updated uses are possible. With the development of cloud computing everything is fast moving and time effective. The serverless architecture of cloud computing ensures it is possible to manage data remotely. So, for big data to move forward it is necessary to transfer storage of big data from traditional hard drives, which are unreliable and often fail and it is difficult to scale them up, to cloud storage services. These cloud services provide nearly unlimited storage with high fault tolerance. So, they are the next step as big data progresses. Unfortunately, for this to happen it will be required for the price of hosting this massive amount of big data on the cloud will have to come down in order for it to be utilised.

“Internet users generate about 2.5 quintillion bytes of data each day” and “In 2012, only 0.5% of all data was analysed” (Petrov, 2019). This shows us we will be generating astronomical amounts of data that in the past was not utilised or processed at all. A short-term solution to be able to process large amounts of data would be, increase the speed of the CPU, storage and network. So, the for the future of big data they will be required to replace traditional algorithms that are not able to pre-process the massive amounts of data. “Highly efficient and scalable data reduction algorithms are required for removing the potentially irrelevant, redundant, noisy and misleading data, and this is one of the most important tasks in Big Data research” (Zhai, Ong, and Tsang 2014).

Data is large and difficult for computers to manage efficiently, so in order to optimise big data management it requires an algorithm that automatically describes big data and relevant processes and generates meta data. Due to the variability of big data it is extremely difficult to generate accurate meta data and traditional database management systems struggle with the scalability of unstructured big data. So once databases can handle big data and generate the large volume of meta data it will be easier for computers to efficiently manage, analyse and visualize big data.

As big data continues to grow and develop, “97.2% of organizations are investing in big data and AI” (Petrov, 2019). As newer technology integrates with the market companies will require people to fill specialized roles as the demand for Big data analytics. This will become a huge part of big data, getting people who are skilled and versed in utilising and processing the data in effective ways that will result in profits for the company.

The future of big data analytics is not set in stone, but it is evident that it is one of the major upcoming thriving industries that will be a behemoth in years to come. For this to happen it needs to be prioritised developed and utilised by industries so that everyone will push to improve the infrastructure around it in order to utilise all the information to the fullest. “Every company has big data in its future and every company will eventually be in the data business.” – By Thomas H. Davenport.

Pearlman, S., n.d. *What Is Big Data and Where Is It Going? - Talend*. [online] Talend Real-Time Open Source Data Integration Software. Available at: <https://www.talend.com/resources/future-big-data/> [Accessed 12 May 2020].

Petrov, C., 2019. 16+ Big Data Statistics - The Information We Generate [2020]. [online] Tech Jury. Available at: <https://techjury.net/stats-about/big-data-statistics/> [Accessed 12 May 2020].

Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu & Fei Hu (2017) Big Data and cloud computing: innovation opportunities and challenges, International Journal of Digital Earth, 10:1, 13-53, DOI: [10.1080/17538947.2016.1239771](https://doi.org/10.1080/17538947.2016.1239771).

Zhai, Y., Y. S. Ong, and I. W. Tsang. 2014. “The Emerging “Big Dimensionality”.” IEEE Computational Intelligence Magazine 9 (3): 14–26.

Tripathi, P., 2018. The Future of Big Data: A Sequential Analysis Of The Disruptive Technology’S Scope. [online] Digital Vidya. Available at: <https://www.digitalvidya.com/blog/big-data-future/> [Accessed 12 May 2020].

Pages by Brian O Connor - 19267827

Big Data and Privacy Protection

As data is now a hugely important factor for companies to use when considering a product launch and the demographic or pricing of that product. The compilation and use of this data bring about its own issues with personal privacy of the people who are the owners of said data. The explosive rate at which the internet has developed has caused people to “leave a lot of traces on the Internet every day” (Zhang, 2018, p. 275) ,and therefore a need for increased security for the data of these people has occurred. Nowadays security is needed in all forms of data as companies use and store this data to understand the present trends of the market. ”How to deal with security and privacy issues in the context of big data is an urgent need for people to have a good solution.” (Zhang, 2018, p. 275).

The current state of big data is such that people don’t trust the security and privacy of the data stored by companies especially with the surge in use of mobile technologies worldwide. “At present, people have serious problems with the security of big data, and think that big data is not safe.” (Zhang, 2018, p. 276). Security issues with big data carry over to personal mobile devices which we carry with us on day by day basis. The loss of these devices means not only loss of data for a company but also loss of data for the consumer that owns the device. The advent of smart homes in recent years has therefore brought more security concerns as these are controlled by the user from their personal mobile devices. The loss or controlling of these personal devices “bring serious security problems to the user’s smart home”. (Zhang, 2018, p. 276).

It is no secret that people are more and more sceptical about how the data they create is used by different organisations and even governments. As such “most Western countries have established special privacy protection agencies to protect citizens’ privacy and information” (Zhang, 2018, p. 277) which allows these countries to create a safe place online for peoples data to gathered, stored and kept from those with malicious intentions. Awareness with regards to data has grown amongst the public over the past few years as people realise the need to protect the private details of their lives over the internet. “The data awareness is mainly directed at the general public and requires citizens to realize the importance of big data”. (Zhang, 2018, p. 277). Awareness of privacy of data must be thought of by citizens in the current world as Zhang states “awareness of privacy solutions in our country needs to be strengthened so that privacy information security can be implemented” (Zhang, 2018, p. 277).

Big data can be used in areas where privacy is an utmost concern such as the healthcare industry. Patient data is highly sensitive information that needs to be protected for the patients in hospitals as well as those who visit general practitioners. Therefore “patient information is stored in data centres with varying levels of security”. (Patil & Seshadri, 2014, p. 763). The cloud and data centres help with storing this crucial data that is used daily to treat patients with various illnesses and injuries. The data also is used to track patients during stays in hospital. The data is in such a way that “Traditional security solutions cannot be directly applied to large and inherently diverse data sets.” (Patil & Seshadri, 2014, p. 764). We are able to see that the data belonging to this industry is often at risk due to the “sophisticated attacks ranging from Distributed Denial of Service (DDoS) to stealthy malware.” (Patil & Seshadri, 2014, p. 764) that plague the healthcare industry annually. Therefore the safety and privacy of the large amount of data that is used in the healthcare industry is paramount especially as attackers begin to use more sophisticated forms of cyber-attacks such as ransomware which requires the payment of a sum of money so they can receive the data of patients back from the attackers.

Data in the healthcare industry and many others needs to be anonymised before it is analysed so that the identity of patients that are in the healthcare system are further protected from cyber-attacks on hospital systems. “Data anonymization prior to analytics could protect patient identity.” (Patil & Seshadri, 2014, p. 764). As the healthcare profession begins to use more and more devices to sort and analyse the data provided by hospitals about patients there is a need to process and analyse said “data in an ad-hoc decentralized manner.” (Patil & Seshadri, 2014, p. 764). The data of patients’ needs to processed in such a way that it is protected and big data analytics allows this to be possible as it can process and analyse the data in an offsite location from the hospitals and then provide the analyses to the doctors and nurse and other medical professionals that need the data to help the patients to recover

The analytics of the data provided by people is required to have their consent due to already in place laws about the use and analyses of data. ““informed consent” from patients is required prior to performing any analytics on patient data”. (Patil & Seshadri, 2014, p. 764). Although laws are in place for the privacy of patient’s data that is then used in big data analytics there is still work to be done to make sure that the laws upholding the need of privacy and security of the data is met. “new laws need to be drafted to clearly illustrate all processes involved in performing big data analytics on patient data”. (Patil & Seshadri, 2014, p. 764). The laws currently in place do not provide enough security during the analysis of private data such as that in the healthcare system and therefore laws have to be made so that big data analytics can be used appropriately in the healthcare industry and other industries where data of people is extremely private.

In conclusion, big data analytics has serious privacy concerns with regards to how it used by industries and how it is analysed by these industries. We must be careful with we use and process the data that is now abundant due to the number of devices that are present in the modern world.

Sources

H. Kupwade Patil and R. Seshadri, "Big Data Security and Privacy Issues in Healthcare," 2014 IEEE International Congress on Big Data, Anchorage, AK, 2014, pp. 762-765, doi: 10.1109/BigData.Congress.2014.112.

Dongpo Zhang,2018/10, Big Data Security and privacy Protection, 8th International Conference on Management and Computer Science (ICMS 2018), Atlantis Press, pp.275-278, Do: https://doi.org/10.2991/icms-18.2018.56,ID:Zhang2018/10

(Page by Luke Bogdanovic; ID-19232691)

The importance of privacy: how Big Data Analytics can be abused by big corporations

Large corporations like Amazon Facebook and Google now have the ability to monitor their userbase in an unprecedented way. Never before could a company gather as much data on their userbase and the effectiveness of their advertising and attempts to sell products. For instance, Facebook can collect data pertaining to the personality, likes and dislikes, and even political views of their users and with the help of big data analytics firms process this information into usable market tactics and business decisions. One high profile case of misuse of data analytics can be seen in the Cambridge Analytica Facebook data scandal “The Times [reported that in 2014 contractors and employees](https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html) of Cambridge Analytica, eager to sell psychological profiles of American voters to political campaigns, acquired the private Facebook data of tens of millions of users — the largest known leak in Facebook history.” – The New York Times[1]. Cambridge Analytica used the data generated from tens of millions of American citizens in order to build up an idea of their political leanings. This Data was then used by interested parties in order to target ad campaigns and misleading news articles in order to sway potential voters  
Cambridge Analytica was again caught up in another scandal involving misuse of Facebook data in 2016 when a former Cambridge employee told the British parliament that Cambridge helped to swing the electorates views towards a withdrawal from the EU ““I now feel sick about what happened,” Mr. Sanni said in his testimony. The officials of Vote Leave “used two 22-year-olds to cheat in the referendum.””- The New York Times[2]. According to this former employee Shahmir Sanni Cambridge used data collected from Facebook, without the consent of the Facebook users in question, to glean the information necessary to help sway the outcome of the UK Brexit vote in favour of leaving by providing lists of people who could have be targeted by political ads and misleading news articles.  
All of this information on the political leanings of the Facebook users whose information had been taken was created by the use of big data analytics, using only the data available on their page “The data Cambridge collected from profiles … included details on users’ identities, friend networks and “likes.” Only a tiny fraction of the users had agreed to release their information to a third party.”- The NY Times[3].  
The Cambridge Analytica case shows how a normal worker in a large corporation such as Facebook or amazon could be told by their superiors to gather useable information from big data analytics without realising that what they were doing was unethical. It is not in their hands what is done with the information they collect or how the data they collect is acquired, it is in the hands of the higher ups in the company, the decision makers. Because of this it is easy for data breaches such as this to go unnoticed by the average joe worker in large firms such as Amazon, Google, and Facebook.   
Large firms such as Amazon and Google use Big Data Analytics to learn about the wants and needs of its users and can use this information to advertise to them items they believe the consumer might need however an issue arises from this because Amazon could use this information if they so desired to promote their own products and hide other companies who sell on their platform from the consumer in order to ensure the consumer buys from them. In this way amazon could put smaller businesses out of business in the guise of providing them with a place to advertise and sell their products. Search for anything on Amazon and their own brand items will be the first to be displayed always if they wanted to they could hide competitors easily.  
Often times with Big Data Analytics companies can know something about their consumers before their consumers know it about themselves. In the early 2000s Target the largest corporation in America began tracking each customers buying habits and extracting useable information from these buying habits such as knowing when to send which customer which coupons in order to get them to shop in Target more. In order to get this usable information Target hired Andrew Pole a statistician in order to figured out customers buying habits. One such item of information he could find out was when a customer was pregnant and and how far along they were in their pregnancy and so on and so forth. This seemed to work for target “Duhigg suggests that Target's gangbusters revenue growth -- $44 billion in 2002, when Pole was hired, to $67 billion in 2010 -- is attributable to Pole's helping the retail giant corner the baby-on-board market, citing company president Gregg Steinhafel boasting to investors about the company’s “heightened focus on items and categories that appeal to specific guest segments such as mom and baby.”” – Forbes [4]. This large increase in profits can show how useful big data analytics can be in business decision making although it can also have unforeseen consequences. In one store in Minneapolis the manager was confronted by an angry father who claimed Target was trying to encourage his young daughter to get pregnant by sending her diaper and crib coupons ““My daughter got this in the mail!” he said. “She’s still in high school, and you’re sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?””- Forbes [4]. In this case his daughter was pregnant without the father knowing it which was quite the surprise for him. Not an unethical use of data but still and unforeseen consequence big data analytics  
The use of Big Data Analytics can be very beneficial for companies but using this information unethically can also be very profitable for them. It is important that these data firms are monitored appropriately and prevented from carrying out unethical business practices in their pursuit of profits. As was the case with Cambridge Analytica with their political use of large unethically sourced Data. In conclusion it is vitally important that the people’s privacy is maintained when using their information for Big Data Analytics

(Page by Jack Ryan ; ID - 19268963)

[1] <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>  
[2] <https://www.nytimes.com/2018/03/24/world/europe/uk-brexit-vote-leave-shahmir-sanni.html>

[3] <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>  
[4] <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#2747890d6668>

Conclusion

In conclusion, we big data is a game changer in how we process data and analyse it. It allows us to gain insights into the different trends there are in the market for products and services. Companies collect data for this reason as well as to know when they should alter prices based on these market trends. The future of big data as previously mentioned is not set in stone and all companies may be able to use this data as an advantage for themselves in the future as well as in the present. The privacy and Security of the data collected by companies and industries is very important as we have a need to protect the people who create the data, this is incredibly evident as our research into the abuse of data that some companies commit shows us that we need more protections for the data people create. Therefore our research into this topic has shown us that the area of big data analytics is a vastly powerful and fast growing part of the computer science and computers industries as it allows lots of users of data to analyse and then adapt things based on this data provided they do so in a way that doesn’t affect the privacy and security of the creators of the data.