THE IMPORTANCE OF SECURITY IN THE AGE OF DATA SCIENCE

Prof. Dr. Mefail Tahiri

University of Tetova, mefailt@gmail.com

Doc.Dr. Ejup Rustemi

University of Tetova, ejup.rustemi@yahoo.com

Abstract

We are living in a world of enormous data, largely known as the age of Big Data. Enterprise data, personal data, social networking, project management, all of these issues generate data continuously. Every aspect of this data needs to be analyzed and managed so we can derive meaning and plan upon them, thus requiring a very sensitive approach to secure them properly. Data security, be it from the side of an enterprise or our own way of management, is crucial to avoid any possible misunderstanding or decision-making based upon breached data which may have been corrupted or manipulated to display information against our interest. This paper will try to evaluate the issue of data security and its impact on our decision at this age which call the Age of Data Science. Security represents an aspect that we encounter almost every day, be it in relation to our personal documentation, our business, our school or any other institution or enterprsie. Data Science is crucial regarding this, that's why we have taken care to use as much adequate material as possible from books written by experts in this field, website articles, scientific papers, statistical data from various institutions, all to clarify our purpose where we want to point and recommend that taking care and securing our data will make our work more successful, something which Data Science is of great help.

Keywords: data, data security, data science, decision-making.

INTRODUCTION

Each time we need to pass to the other side of the road, we should watch right and left so we can be safe enough to make the first step, otherwise, we may put ourselves in danger by being run down by a car or a truck. The same principles apply when we deal with valuable data which can be crucial for a given organization no matter the nature of its businesses. It can be a car manufacturing company where data can be crucial for the safety of the driver and passengers; it can be a medical facility where data analysis can make or break the company – If data is used properly it will make for a great medical product, if not it can result in fatal consequences. The point is that data is vital in any field imaginable, that's why keeping them safe and out of the reach of those who don't know how to act with them, or those who simply want to harm, is a very sensitive aspect of data management and analytics. This is where Data Science and security intervene to make way for successful progress.

ARE WE SAFE?

We are certainly leaving in the Age of the Internet, where almost anything we do is displayed online, where it can be seen by those who are close to us. But this is not always the case; sometimes, especially when we are unaware that we have exposed ourselves to somebody who will try to use the data about us to harm us.
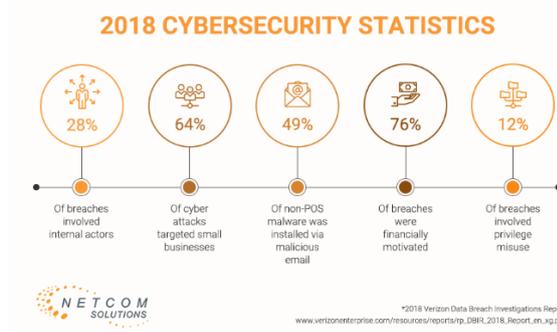
Fig. 1 2018 Cybersecurity Statistics (https://www.netcomsolutions.net/)

With the advancement of social networks, things have gotten even more complicated. Platforms such as Facebook, Instagram, Twitter and other similar, while can be very helpful for communication and sharing, still, they can also represent a door through which our privacy may become breached by others. But if we dwell even deeper into this issue, we will see that users of such platforms cannot be harmed just from anonymous hackers who at times do harm just for fun, but from the platforms themselves. The Cambridge Analytica scandal that hit Facebook is a great example of such an issue. From the scandal, we learned that Facebook has sold private data so others can use for fake news to help political campaigns.
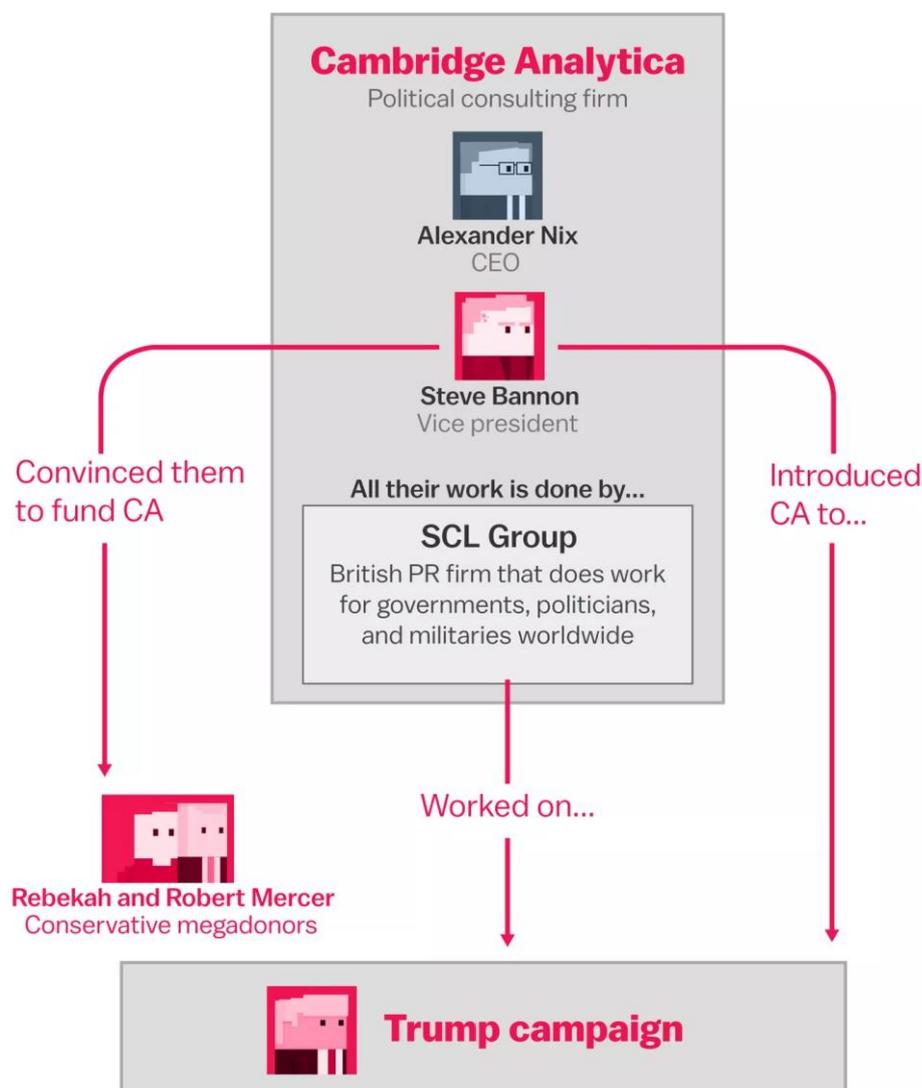
Fig. 2 Political Scheme of Cambridge Analytica (https://www.vox.com/)

Mobile devices have aggravated this issue to a certain degree. Previously, when people were mainly using desktop devices to access the Internet or to communicate with others, devices such as smartphones have completely changed the communication landscape, in one way facilitating enormously how we cooperate and exchange, while on the other, exposing us to the world where there are people who are waiting to attack our privacy.

Fortunately, if we are aware of such things, we can always use software and a lot of time just common sense to avoid any such inconvenience.

DATA IS GOLD

Raw data is like finding a goldmine that will need to be processed to make it very valuable. Data is a true representation of who we are, what we do, how we think. Analyzing the data will generate information, which in return will generate knowledge, all of which can lead to machine learning systems which can provide a helping hand to humans.

Every time we put data into a system, we provide valuable assets that if processed accordingly will benefit ourselves and the company we are working for, or are owners of.

Considering this importance, we can easily understand that keeping that data safe is of utmost importance, because if it gets in the hands of those who mean us harm we may face large scale problems, even scandals. This should be particularly underlined when things may reach the competition. Enterprises or any other organization are always on the move and in constant search for how to win their competing organizations, that is why the majority of times they keep a lot of information in secrecy. If that information is not secure enough it can be exposed and other organizations may use it for blackmail or to change their strategy to be more competitive against others. This may sound like something out of a movie, but it is something that is constantly happening in the business landscape.

Enterprises analysing big data from any data source, 2018 (% of enterprises)
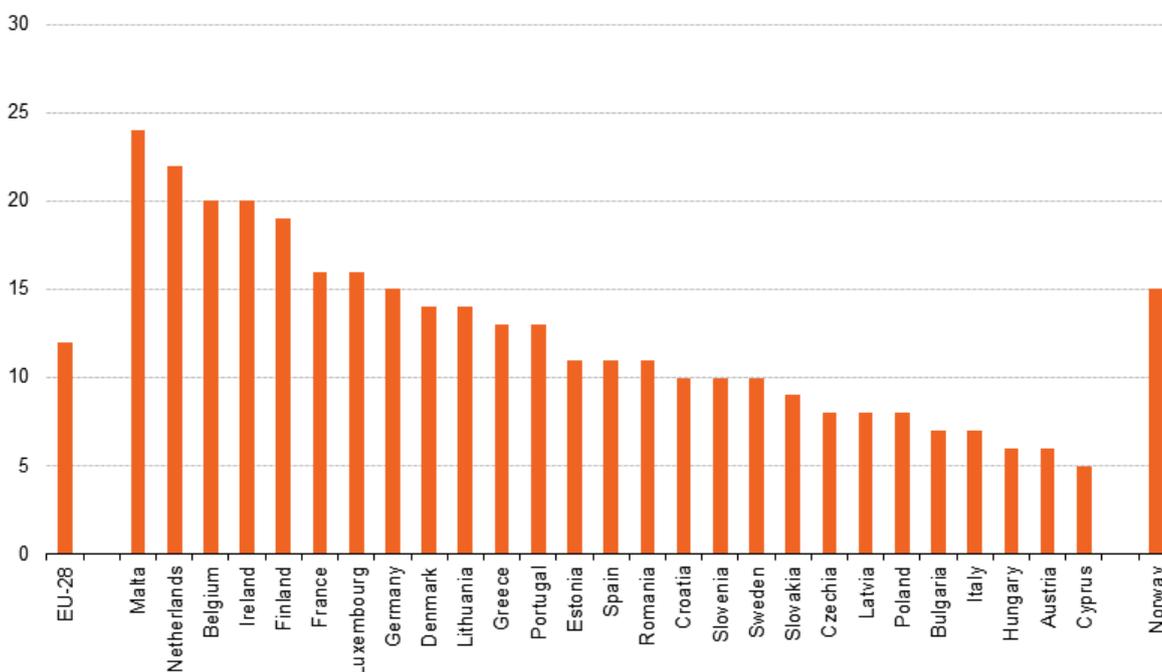
EU-28 without United Kingdom. United Kingdom, Iceland: no data available
Source: Eurostat (online data code: isoc_eb_bd)

Fig. 3 Enterprise Analysing Big Data (https://ec.europa.eu)

Even the largest IT companies use data from customers to attack their direct competition. In recent years we have seen many attacks between companies such as Microsoft, Apple, Samsung and many others. These companies have used data from customers who were unsatisfied with the services or products of the other company to gain the attention of those customers and to possibly direct them towards their own products. Apple, for example, used the case when Samsung Galaxy Notes 7 products were exploding, Microsoft used the dissatisfaction of customers with Apple for making faulty keyboards on their MacBook Pro line of products or using only limited colour options for their iPhone 5 devices.

Whatever the case Data Science should provide the tools to maintain any amount of data so an organization can achieve its goals and maintain a solidified integrity.

A data scientist should not only analyse and evaluate data in a given company, but they should also manage the data gathered by companies who make statistics of the severity of data breaches, data theft and anything similar that directly offends human privacy.

DEFENDING OURSELVES

So the question now arises, how we can defend ourselves in all this confusion? Or a better question, whom we can trust? To answer these questions is not an easy task, especially if we have in mind that there are many parts involved, all of which think and act as if they are right and all others are wrong, or purportedly act against their interests. Data Science, in building an analytic environment build upon data extraction, comparison, analysis, statistic, history of changes in a particular environment or specific organization,

etc.; in other words, it actualizes every aspect of data management, so we can better understand how everything relates and affects an organization

Fig. 4 Defence Scheme (https://geeks.co.uk)

CONCLUSION

I think that it is quite easy to conclude that data security is crucial to the proper function and development of organizations around the world. It is also easy, especially if we read and understand the statistics properly, to conclude that people are prone to give away information about themselves without thinking about their privacy and wellbeing. Things get even more complicated if data breach affects large organizations where we may witness losing millions or even billions of dollars in assets and value. The best way to be safe, I believe, is not just to use antivirus, antimalware or any other software, but to be aware of the danger and not to expose data when there is no need to do so. This is where Data Science should provide a helping hand to individuals and particularly organizations, will do their work in a safer environment, and in that way, they will open the doors of the many possibilities for a brighter future.

REFERENCES AND BIBLIOGRAPHY

Davidoff, Sh. (2020); *Data Breaches Crisis and Opportunity,* Addison-Wesley, Boston

Ozdemir, S. (2016); *Principles of Data Science,* Packt Publishing, Birmingham

Prasad, R., Rohokale, V. (2020); *Cyber Security: The Lifeline of Information and Communication Technology,* Springer, Cham

Hoffer, Ramesh, Topi (2011); *Modern Database Management,* Pearson, Boston,

Jurney, R. (2017); *Agile Data Science,* O'Reilly, Cambridge,

Laursen, Thorlund, (2010); *Business Analytics for Managers,* Wiley, Hoboken

Patil, DJ. (2011); *Building Data Science Teams,* O'Reilly, Cambridge

Peng, R.D., Matsui, E. (2015); *The Art of Data Science,* Leanpub, Victoria

Vermaulen, A.F. (2018); *Practical Data Science: A Guide to Building the Technology Stack for Turning Data Lakes into Business Assets,* Apress, West Kilbride North Ayrshire

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:V1_Enterprises_analysing_big_data_from_any_data_source,_2018_(%25_of_enterprises).png

https://geeks.co.uk/2017/08/how-to-defend-your-business-against-cyber-attack

https://www.netcomsolutions.net/2019/03/data-breach-statistics-in-2018-that-smbs-should-know/

https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram