LUISS Guido Carli
LIBERA UNIVERSITA' INTERNAZIONALE DEGLI STUDI SOCIALI

Dipartimento: Economia e Finanza Cattedra: Law and Economics

Nowcasting:

How Big Data predict the present

RELATORE

Prof. Andrea Renda

CANDIDATO

Riccardo Dinale

168631

2013-2014

Abstract

This paper aims at explaining how Big Data can help society and businesses to predict real-time changes in the external environment, in order to increase flexibility and efficiency. Big Data are defined as high volume, velocity, variety, and veracity information assets. An analysis of the most recent theories and studies about Big Data and nowcasting leads to the Predpol case - where Jeff Brantingham created a nowcasting algorithm to fasten police intervention in Los Angeles - which is compared with IBM Blue CRUSH technology - a similar nowcasting predictive tool for crime fight. A second case study shows how nowcasting can help at predicting the mood of a nation. The paper shows that nowcasting is an increasingly important econometric technique for predicting events in real time by using Big Data, which will benefit social policing and enterprises. This outcome changes the way statistical predictions have always been used and it opens new huge opportunities for societal issues and business profitability.

Table of Contents

Introduction

In the first chapter, the study provides the full Big Data context: how it is defined; how is the industry performing; major upsides, as theory testing or improved decision making; and important downsides, as scalability and refresh costs or privacy issues. The analysis is undertaken in order to create a comfortable framework to the reader by reviewing theories and thoughts about the new technology of Big Data.

The goal of this paper is to give a brief introduction about Big Data research in order to better understand an exciting econometric technique labeled as “nowcasting”, which will be described in the second chapter. The theoretical framework fails to show how Big Data can be practically advantageous to society, and this is why the technique will be analyzed through the use of three case studies. In the second chapter I will first define and analyze the technique, giving a proper understanding of the subject in order; secondly, a selection of case studies will be shown in order to underline the many possible and practical uses of an efficient Database Management System (DBMS) using nowcasting methods. Firstly, Predpol and IBM Blue CRUSH, which are discussed and compared based on their effects on crime rates, respectively in Los Angeles Foothill and Memphis. As we will see, both show stunning and comparable performances, thereby confirming real-time predictions potential. The last case describes how the mood of the nation can be nowcasted by using social networks data and “words clouds”.

Chapter 1: Big Data framework

Big Data is an increasingly important technology, and scholars around the world are trying to identify its strengths and weaknesses, harnessing what are the real advantages of running a Database Management System (DBMS) and more importantly investing on it. This chapter will provide a detailed overview of such technology.

1.1 History and Definition of Big Data

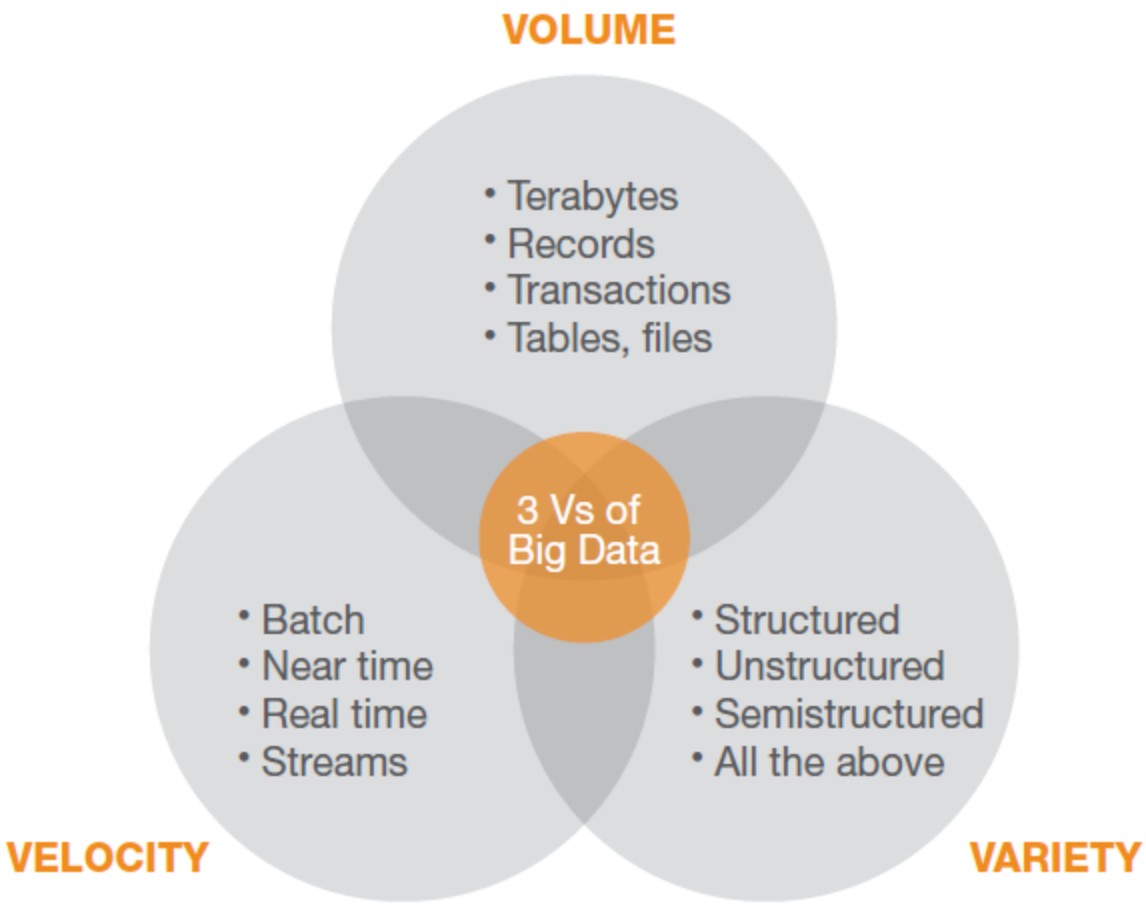
The history of Big Data starts in the early 2000s: data volumes were increasing at levels that hardware could not process and analyze (Russom, 2011) . The firms' IT departments were struggling to find the necessary resources to manage and exploit such massive databases. This data scalability issue was naturally solved by technological advancement in raw processing power - CPUs, RAM - which turned a huge problem into one of the greatest opportunities firms could seek.

Big Data were roughly defined first by Laney (2001), when data growth challenges were found to show three dimensions: Volume, Velocity, and Variety, respectively the amount of data, the physical bandwidth and protocol, and the range of data types.
This definition was renewed in 2012 (Laney), when the three Vs were integrated in a comprehensive frame:

"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."

High Volume, Velocity and Variety information assets are hence defined as Big Data.

Figure 1 "The three Vs"



Source: Russom, 2011

The above figure (Figure 1) clearly explains the related characteristics of each of the three dimensions that describe Big Data. In detail, Volume is represented by the amount of data stored, for instance records, transactions or tables, and quantified in Terabytes. Further, Velocity regards the refresh rate of information - near time, real time - as well as the actual stream of data and its flows speed. Finally, Variety includes the range of data types, namely, whether it is structured, unstructured or semistructured

Another V, Veracity, is being added by an increasing number of scholars, as Buhl et al. (2013), Mattman (2013), Gattiker et al. (2012). It is defined as the meaningfulness of data, which has to be trustworthy and reliable in order to bring enhancements in IT management and decision making.

Big Data's functioning is well-described by Boyd et al. (2011):

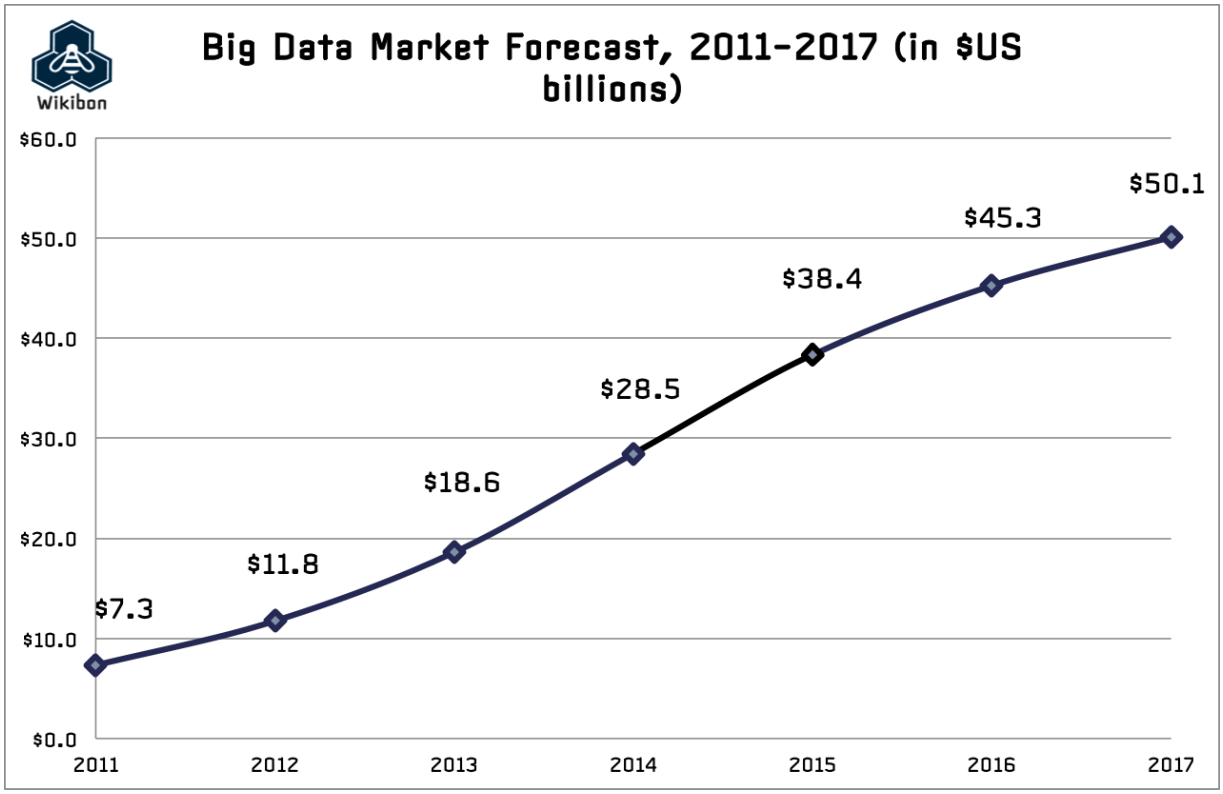
> *"Due to efforts to mine and aggregate data, Big Data is fundamentally networked. Its value comes from the patterns that can be derived by making connections between pieces of data, about an individual, about individuals in relation to others, about groups of people, or simply about the structure of information itself."*

The potential for discovering patterns in the underlying data is what declared their success in the business world, and what is making firms thriving. The top seventy Big Data vendors, 18.6 billion dollars as a whole, enjoyed 58% of annual growth in revenues, composed mostly by services and hardware sales (Gil Press, 2014). The first pure-play vendor,

Palantir, earned 418 million dollars in 2013, growing by an astonishing 118% in one year (pure-player as revenues are totally derived from Big Data products and services).
Loosening the constraint of pure-play companies, the most important Big Data vendor is IBM, earning almost 1.4 billion dollars in 2013. An IBM case of Big Data and nowcasting will be covered in the second chapter, showing the technology social benefits. Following IBM, we find HP and Dell, respectively gaining 869 and 652 million dollars.

Obviously, the market is still expected to grow at a considerable pace: according to Kelly (2014), the aggregate revenues will increase from 18.6 to 50.1 billion dollars in 2017. The growth is driven by an improved articulation of vendors' offerings and the increased number of partnerships, which will lead to increased adoption in the mass market, because of the technology's increasing maturity, and in the enterprise market, caused by important steps toward privacy and security of Big Data products and services (Kelly, 2014).

Figure 2 "Big Data market forecast"



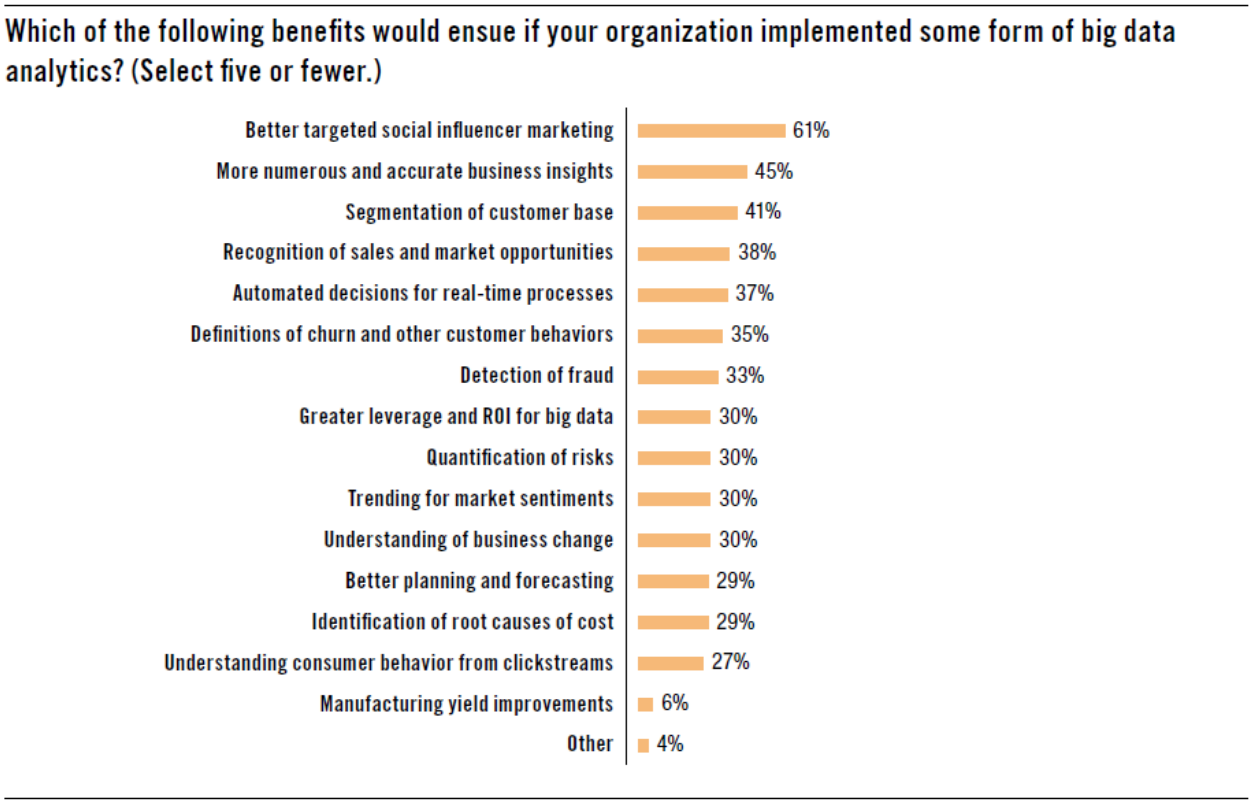
Source: Wikibon, 2014

1.2 Advantages and potential of Big Data

Big Data are considered almost universally a new, exciting technology. According to McGuire et al. (2012), all companies should analyze whether Big Data can increase their competitive advantage, since a refusal to upgrade the business IT could lead to serious losses. The author shows many ways to leverage Big Data technology: first, information becomes increasingly transparent, collecting data from traditional sources and reducing the inefficiency of transfer among users; second, this continuous increase in data

volume is directly and positively correlated with the possibility to analyze every business area, fostering efficient decision making and product designing, which leads to the last point, i.e. narrower segmentation, tailored products and a new start for next-generation products and services. The last remarkable topic is the use of Big Data to evaluate real-time theories and business decisions, in order to obtain a fairly reliable simulation of the desired strategies. The advantages are confirmed by TDWI's survey (2011), which asked to 325 firms of different sizes which kind of benefits would they receive from Big Data implementation.

Figure 3 "Benefits from Big Data analytics"



Source: Russom, 2011

While the most popular answer belongs to the social marketing area, the following three answers confirms the chance for improved decision making, accurate segmentation, and new products or services as described by McGuire et al. (2012). It is interesting to notice how firms do not rely on Big Data analytics for "manufacturing yield improvement" purposes. This scenario opens new unique opportunities for strategic managers.

McGuire et al.'s (2012) view is shared by McAfee et al. (2012), which stresses the importance of Big Data for knowledge management and decision making purposes. The authors show empirical facts, obtained by cooperating with MIT colleagues and McKinsey's specialists, about how this technology can boost enterprises performance. The analysis was based on structured interviews of top executives and collection of performance data. The outcomes are clear:

> *"Companies in the top third of the industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors. This performance difference remained robust after accounting for the contributions of labor, capital, purchased services, and traditional IT investment. It was statistically significant and economically important, and was reflected in measurable increases of stock market valuations."*[1]

[1] McAfee et al. 2012

The article also underlines how big data can reduce a certain kind of decision making process, called “HiPPO”: The “Highest-Paid Person Opinion”. Many companies rely on top executives whose instinct is supposed to drive the most important decisions for the business. The real issue is when companies over-rely on such individuals, adopting a suboptimal decision making process. Big Data technology can improve it, by mixing expert businessmen instinct and analytical results.

Finally, McAfee (2012) lists 5 challenges for effective implementation:

1. “Leadership”: big Data do not substitute a capable, longsighted leader team. Successful companies still need clear goals and visions, and most of all human leadership creativity.

2. “Talent Management”: as the importance of data analysis grows, the need for data scientists talent increases exponentially. Big Data’ technology is not sufficient: enterprises must seek for data managers and experts able to create value from the implementation.

3. “Technology”: discrete investments on new infrastructures and commodities must be considered as central in big data strategy.

4. “Decision Making”: the right data and information must be crossed with the right talent and skills in order to increase functionality and exploit the analytical results.

5. “Company Culture”: a more general challenge, as companies need to effectively understand what their culture is, and what do they know.

These five challenges are risks and opportunities at the same time.

But Big Data can also be analyzed by their practical effects on society, as Tene et al. (2012) describes. He found seven areas where Big Data brought new, clever advantages:

1. Healthcare: by crossing Adverse Event Reporting System (AERS) data and Microsoft’s search engine Bing, researchers were able to discover diabetes-inducing side effects caused by the mix of two different drugs. This fact underlines how medical records will be the key for future analysis and improvements.

2. Mobile: scientists are analyzing mobile communication in developing countries, in order to predict crime, food shortages, and to improve schooling outcomes.

3. “Smart Grid”: this term refers to a “bi-directional flow of information and electricity” (Tene et al. , 2012), instead of the most common infrastructures. It would allow service providers to collect data, improve efficiency and quickly locate power

issues, while at the same time enabling customers to be better informed about their electricity consumption.

4. Traffic Management: governments could issues far more accurate policies in order to best fit the actual cars usage by drivers, while the latter could exploit real-time traffic information. Finally, urban planners would have more powerful tools for enhanced decision making.

5. Retail: customers profiling enables businesses to consistently advertise the right customers. Profiling and behavioral information are valuable, while the identity rises privacy issues, which will be discussed in the following paragraph.

6. Payments: Big Data can help merchants to detect fraudulent card payments by using buyer purchasing history.

7. Online: the new, powerful social media are facing an exponential increase in data. These IT companies are leveraging such databases in order to maximize efficiency and understand their customers' needs.

The "mobile" area will be further analyzed in Chapter 2, where two bright cases of crime prediction through use of nowcasting and Big Data are analyzed.

1.3 Disadvantages, ambiguities, and policy issues of Big Data

The concerns about Big Data are never-ending. While small businesses are evaluating such opportunity, big enterprises are already practicing it, creating a bandwagon effect on competitors. Some scholars depict a quite different vision of this new technology, since as advantages become clearer, the same disadvantages do.

According to Boyd et al., Big Data, while being potentially exceptional for nearly every field of research, leads scholars to practice apophenia, defined as seeing patterns where they do not exist. This would be caused by the immense amount of data available, whose analysis creates links and relations among single elements where actually none exists. Apophenia definitely voids any result, thereby calling for careful use of Big Data and their statistical outcomes. Moreover, Boyd states that Big Data face many issues, especially ethical concerns and a new concept of digital divide, where areas with access to Big Data outperform all the others. He also critics the data itself, since bigger data is not necessarily of higher quality; finally, data is not always interchangeable and comparable among different sources.

Some scholars state that Big Data definitely show dangerous flaws for researchers. Richards et al. (2013) list three important paradoxes of this technology:

1. “The Transparency Paradox”: while Big Data are supposed to make information transparent, the process of collecting data is the true opposite as it shows many

layers of legal, physical, and private property barriers. The author adds that the growth of mobile devices and sensors will exponentially increase both the amount of data produced and the need for a transparent collection process. This analysis is in contrast with the first point of McGuire et al. ‘s (2012) argument, as cited above. In fact, the author states that Big Data will increase transparency and transfer of data, while they will enable firms to exploit the growing data volume for business decision making and products tailoring. Data and transparency are therefore evaluated as both risks and opportunities.

2. “The Identity Paradox”: as Big Data are used to learn more about customers i.e. us, companies could be able to influence people’s choices, hence undermining the individual thoughts.

3. “The Power Paradox”: while the technology seems to help ordinary people, Big Data owners are different entities and institutions. The author states that this new trend will create winners and losers, and chances are that ordinary people will belong to the second category.

Buhl et al. (2013) notes that Big Data are not adopted by the majority of the potentially interested companies because of privacy concerns. The collection of data is not frictionless with respect to international laws. This topic is becoming increasingly important as IT companies grow and develop bigger databases and new analytical techniques.

Privacy importance is proven also by the NSA scandal, started on June the 6th 2013 (Lütticke, 2013), when Edward Snowden, employee at a consulting firm, revealed to *The Guardian* how the NSA was spying millions of people by exploiting IT servers and phone calls records.

It is not surprising that many scholars are focused on the privacy side of Big Data. Tene et al. (2013) describes the "incremental effect" of data collection: as private information stemming from an individual grows, it becomes increasingly harder to detach those data from the source, causing people profiling. In fact, the author asserts that "once data are linked to a single individual, they become more and more difficult to disentangle", citing an experiment that managed to link real users with unidentified data from Netflix by using additional publicly available information.

Privacy is becoming more and more sensitive also for policymakers. The EU decided to reform the Data Protection Directive in January 2012 in order to cope with new technologies and needs. The reform aims at reducing legislative fragmentation among states as well as simplification of existing rules. Some key changes are the importance of responsibility and accountability of companies with regard to data protection, the creation of a single national supervising authority, easy access to personal data by customers, the "right to be forgotten" - the possibility to delete all personal data, and increase in police and judicial coordination (European Commission, 2012). The European Commission states that the proposal will allow 2.3 billion euros of savings, achieved by reducing administrative

efforts and unnecessary legislation. Moreover, the reform should reinforce consumer confidence in online services, becoming an important growth driver.

However, the debate is controversial, as scholars provide different interpretations of the proposal's outcomes. According to Rubinstein (2012), the efforts are not going to succeed, since the new proposal shows at least three major flaws: firstly, the DPD relies on consumers' informed choice, while empirical studies showed that consent is always given, since the privacy policies are either not read or not understood by customers; secondly, the DPD asks for data minimization, the enforcement of which is definitely uncommon; finally, the general fit of DPD is outdated, as it fails to recognize the many ways people share and manage their own data.

Chapter 2

2.1 Nowcasting: a new, powerful technique

Data analysis has always been affected by the frequency and the delay of data releases, which allowed low frequency predictions and estimates. At least until the last years, when a new technique called “nowcasting” emerged in the economics research field. The term refers to its most important feature: the ability to predict a real time event. Banbura et al. (2010) states a broad definition:

> *“We define nowcasting as the prediction of the present, the very near future and the very recent past”*

According to the British weather forecaster “Met Office” (2011), the term “nowcasting” was coined in 1980s by Met Office scientist Keith Browning. It referred to very short-range forecasts stemming from a series of radar images

As a matter of fact, this technique was commonly used for meteorological purposes in the last decades, allowing scientists to predict thunderstorms (Mass, 2011). It was sensibly limited by the data source, meteorological radars, which allowed only short-term computations. The change from steady state assumption methods to variables temporal change models signaled the advent of new, accurate predictions (Dixon, 1993). In the last

two decades radar-based models and the quality of mesoscale data enabled weather scientists to accurately predict precipitation levels and thunderstorms (Mass, 2011).

It is interesting to notice how an important tool for weather scientists became a sound opportunity for both society and business. As cited in the introduction, the improvements in data management systems and the lowering of storage costs now allows to nowcast macro and microeconomic variables. This marks a neat difference with forecasting, since the real-time technique allows a fairly higher frequency of predictions, which result in higher accuracy. Still, nowcasting needs contemporaneous disaggregated data, and it therefore faces 4 main issues when predicting aggregate variables - i.e. GDP - (Castle, 2013):

1. *"Missing-data problem":* not all the data are known before the actual nowcast. This requires the use of proxies or constructed variables.

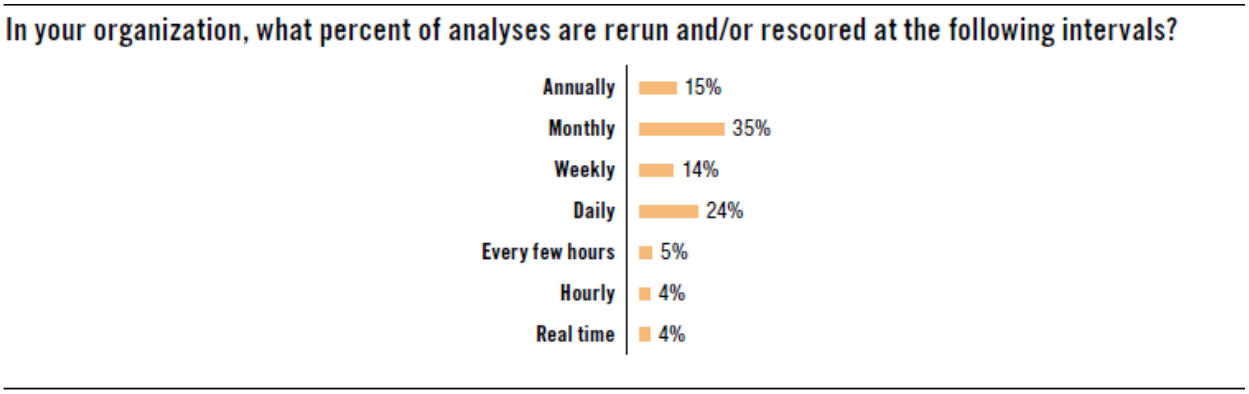
2. *"Latency problem":* the frequency of predictions often results higher than data releases' one, therefore creating a lag.

3. *"Changing database problem":* data is released and updated at a certain frequency, but every source does so at a different time, as a consequence of the previous issue.

4. *"Measurement error problem":* even when the first three flaws are fixed, disaggregated data still face the possibility of being revised during or immediately after the whole process.

These difficulties explain why the majority of organizations do not adopt real-time techniques (Russom, 2011). In fact, only 4% of 96 respondents, composed by organizations using database management systems integrated with Big Data, claimed to refresh the data content in real-time. The majority prefer a far higher standard, mostly daily or monthly. The reason is straightforward, as high frequency increases data volumes.

Figure 4 "Refresh times and usage"



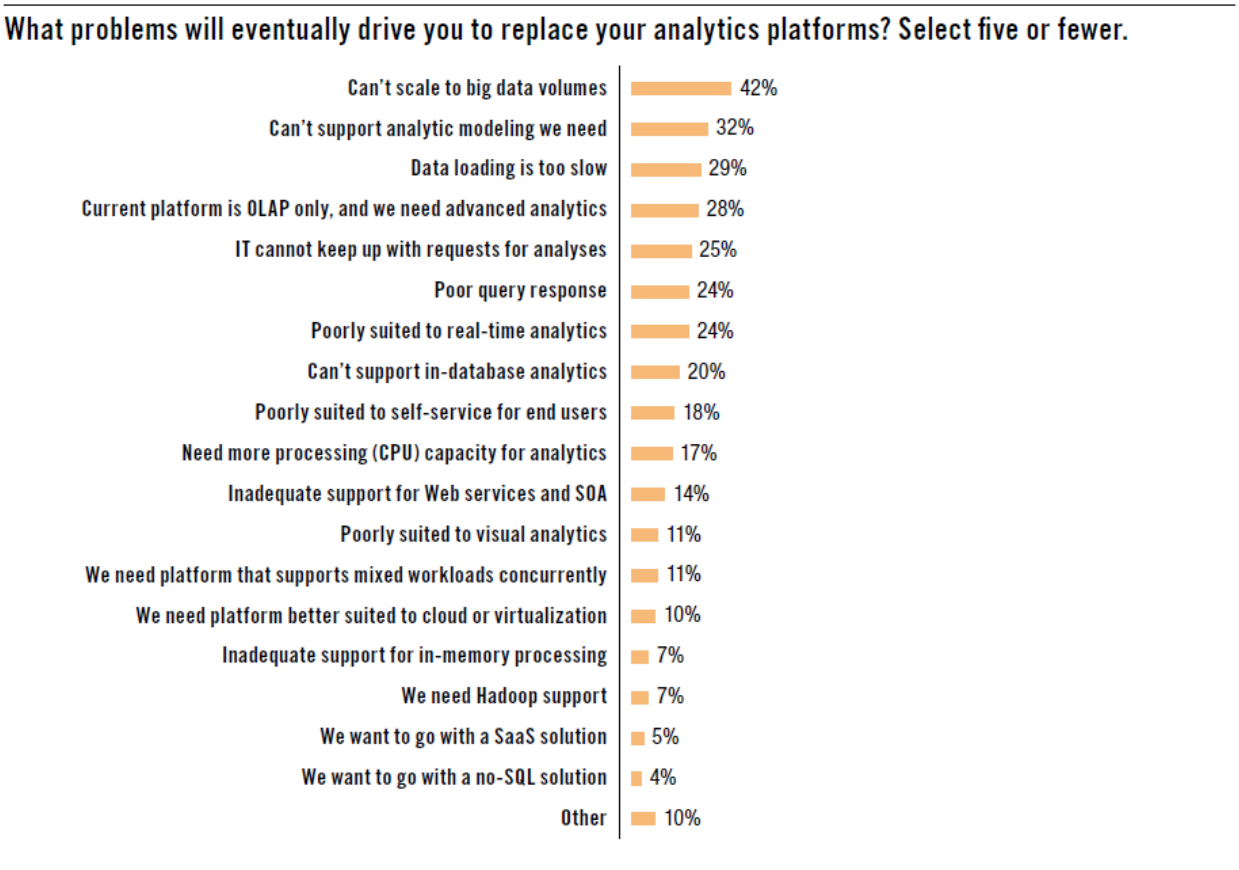
Source: Russom, 2011

This is neatly linked to the surveyees' most important drive for replacement: the impossibility to scale to Big Data volumes, as shown by Figure 5. The surveyees were asked to state the drivers of their analytics platforms future replacement. The correlation

between low data refresh frequencies and volume scaling is clear, since real-time refresh rates need high Velocity, which causes higher stream of data and consequently increasing data volumes. Scaling needs investment, which explains why refresh rates are limited for budget limited firms. It can also be noticed that the third most popular answer was “Data loading is too slow”, which entails the whole definition of Big Data as a three Vs asset.

Still, real-time tools are considered the main trend for Business Intelligence sectors, and they therefore have potential for considerable growth.

Figure 5 “What leads to platforms replacement”



Source: Russom, 2011

2.2 Case Studies

It has been so far discussed the essence of Big Data and the mechanism of nowcasting techniques, but how do they actually merge? In this section, I will explore some important case studies that succeed at exploiting the technologies' potential. Moreover, the chapter aims at framing the advantages of such interaction for policy-making purposes.

2.2.a "Predpol"

A striking example of societal advantages is a company called Predpol. As retrieved from its website, the firm's mission is simple:

> "*Place officers at the right time and location to give them the best chance of preventing crime."* [2]

A simple and clear commitment to crime, which is arguably one of the worst societal activities. In fact, the crime sector generates damages for a significant percentage of countries GDP, as in the US case, in which it consumes around 2 percent of the gross domestic product (Chalfin, 2013). This statistic is even more striking when compared to the education sector's revenues, around 1 percent of GDP. The expensiveness of crime is confirmed by official US data: in 2003 criminal activities represented a burden of

[2] www.predpol.com

approximately 180 billion dollars for State, locals and federal entities (Bureau of Justice Statistics, 2003).

This is the market opportunity Predpol creators found out, an inefficient and expensive gap in society. The project started as a collaboration between UCLA, Santa Clara university, and UC Irvine, where social scientists and mathematicians had been working on algorithms able to analyze historical data as well as be able to adaptively learn and apply advanced mathematical modeling. The project was conducted alongside crime analysts and officers from Los Angeles and Santa Cruz police departments. As a matter of fact, the software has its major successes in Los Angeles Foothill division.

But how does it actually work? The system was inspired by storms nowcasting, and it predicts crimes the same way: officers are given a 500 feet times 500 feet prediction box map that highlights an hotspot for a close-to-happen crime, increasing the police attention and resources on that area. The software does not show a result based on past crimes, but it actually tries to predict the future, precise criminal activity. Because of its own nature, Predpol does not require any particular hardware or training: officers can access the predictions through their daily tech device whenever it is necessary, while managers do not have to invest a higher budget on accessories or complementary assets.

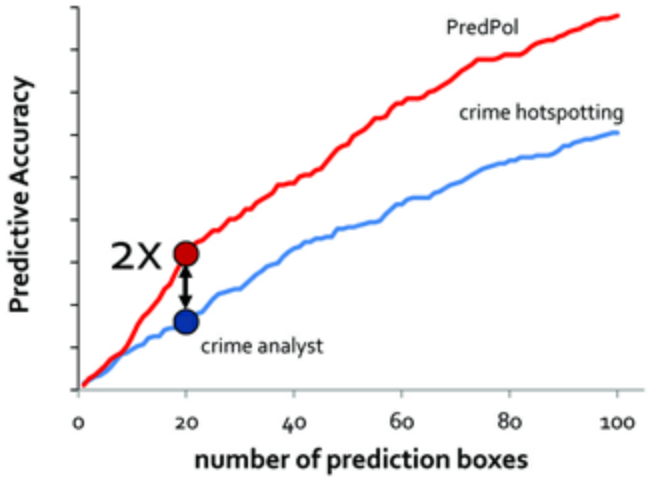
It could be argued that Predpol breaches privacy laws in his attempt at exploiting a Big Data of criminal activities and urban patterns, but the company does not record any individual data about offenders and victims.

Finally, the company adopted the annual subscription business model, which enables the software to reliably run and predict the whole year.

But what makes Predpol outstanding is not the impressive list of features, as the accurate 500x500 feet prediction boxes, the machine learning or the mobile access for officers. It is the actual and consequent impact on crime. The company is proud of the following empirical results in the police departments where the service is currently running:

1. *Prediction accuracy*: the software has proven to be approximately twice as accurate as crime analysts. The result is robust since the pattern does not change with a marginal increase in the number of prediction boxes, as shown in Figure 4.

Figure 6 "Predpol's prediction accuracy"



Source: http://www.predpol.com/results/

2. *Impact on crime rate:* the impressive accuracy would not suffice in case of unchanged crime rates. But the higher reactivity of Foothill officers and reliability of

prediction boxes had an interesting effect on crime: after only 4 months, crime decreased by 13%. The numbers are fascinating, since the city's departments where Predpol was not operating faced, during the same time interval, an increase of 0.4% in the number of crimes. Moreover, as we can infer from the chart below, in one year crime decreased by even almost 30% in the spring time period.

Figure 7 "Crimes per day decrease"

Source: http://www.predpol.com/results/

The Foothill division faced an historical event because of Predpol: the first day without any reported crime (Predpol Blog, 2014).

This case study is a strong advertise for the so-called "predictive policing". Predpol, by joining standard (officers, human resources) and innovative tools, enabled a huge

change in urban behavior. This confirms one of the most important advantages of the nowcasting technique: the potential impact on society. It also demonstrates that resource efficiency is still a distant achievement for modern societies.

2.2.a.1 "Predpol" comparison: "Blue CRUSH"

Predpol achieved outstanding results, which however could be contingent and tied to the context. In order to understand if crime prediction is potentially universally effective, I will discuss a similar case study: Blue CRUSH, or Blue Criminal Reduction Utilizing Statistical History.

Similarly to the Predpol case, the partnership between Memphis PD and Blue CRUSH started when the Larry Godwin, Memphis' Director of Police Services, met Dr. Richard Janikowski, professor of Criminology (IBM, 2011). The latter proposed a solution aimed at optimizing police resources and timing, a technology focused on finding the "hot spots" of crime activities. Godwin asked Richard to integrate police officers and patrols' expertise with the predicting software immediately after a successful operation directed by analytical predictions, in order to gain a predictive tool augmented with officers' knowledge and instinct. It was a real breakthrough, since the MPD needed 500 more agents for facing crime growth, a program that, unfortunately, would have required 6 years. The deal was struck between the major and Godwin, creating the project called "Blue CRUSH".

After 6 years, its empirical results are as amazing as Predpol's ones:

1. An average of 30% crime reduction, with a 36.8% at the highest peak.
2. Violent crime reduced by 15%
3. Share of cases solved by MPD increased by 400%, from 16% to 70%
4. In general, improvement at allocating resources with budget limits.

The MPD now organizes weekly analysis of successful and unsuccessful operations to understand what failed, which took the name "TRAC" as "Tracking for Responsibility, Accountability, and Credibility". Moreover, Godwin himself congratulated with Blue CRUSH for such results, explaining that they are now able to shift resources more easily, but, most importantly, optimizing effectiveness.
The key to success was accountability, which was achieved through standard templates for officers reporting, in order to avoid obscure results.

Blue CRUSH case study confirms the previous outcomes: nowcasted predictive policing supported by Big Data is a viable and advantageous resource for social issues such as crime. Their application to other sectors is potentially beneficial, as quoted in the first chapter, but it still lacks of hard data compared to the case studies.

2.2.b: Nowcasting a nation's mood

The era of social media has started, and what are they if not Big Data themselves? Facebook and Twitter count millions of users, respectively 1.28 billion and 255 million (Smith, 2014), and every single user creates data about feelings, locations, preferences, networks at an incredible pace. In 2010 (Lansdall-Welfare et al., 2012) a team exploited such data flows for a nowcasting experiment: is it possible to predict the incidence of flu by analyzing tweets content? The team's machine learning algorithm found the key words identifying the common illness by maximizing the correlation with the mapping words-flu, which happened through collection of actual flu levels from the Health Protection Agency (HPA). The algorithm result was a "word cloud", a collection of terms weighted for their linear effect on flu levels. The result was positive: flu levels can be approximately nowcasted by using Twitter content.

The team felt so encouraged that decided to forecast the mood of the nation using a similar method. They analyzed tweets from 54 UK cities during a 30-months interval, which accounted for a total of 484 million tweets generated by around 9 million different users.

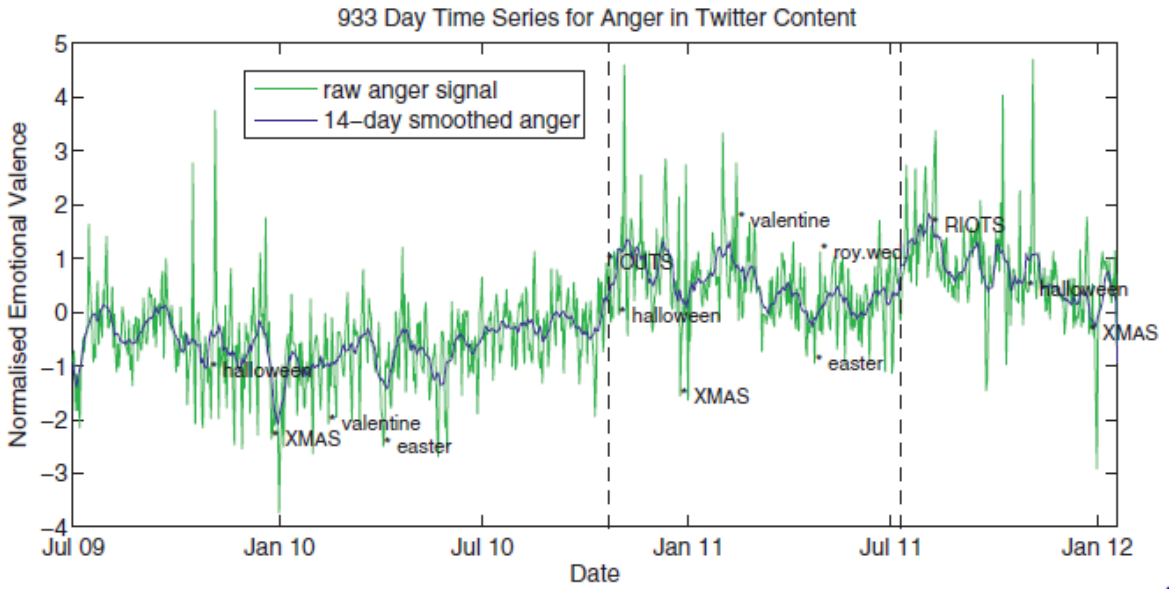
The actual mood is predicted through the so-called "citation-sentiment analysis" (Lansdall-Welfare et al., 2012):

"Each of the basic emotions (fear, joy, anger, sadness) is associated with a list of words, generated by a combination of manual and automatic methods, and successively benchmarked on a test set."

First, some sanity checks were made: the algorithm found peaks of joy and fear respectively during Christmas and Halloween, because of a higher count of keywords during those periods. This results clarifies the need for a more sophisticated software able to filter common expressions - i.e "happy new year" - in order to reduce the outcome bias.

An interesting result was related to the date October 20th, 2010: the Prime Minister Gordon Brown announced wide spending cuts. The algorithm found an impressive peak in anger (Fig. 3) and, in general, a more negative mood, whose shift was still visible in the next year predictions.

Figure 8 "Anger predictions"



(Lansdall-Welfare et al., 2012)

Moreover, the software matched the summer riots of August 2011 with a high level of anger. What is fascinating though, is that such mood gradually increased in the previous months. It follows that such events could be somewhat predicted, even if correlating historical events and data result *ex-post* is by no means similar to their nowcast in real time.

This case study shows that data-driven methods can be exploited in social sciences, even if requiring polishing, careful interpretation, and nontraditional data sources.

Conclusions

Big Data integration is becoming increasingly important for businesses as the complementary technology drops in costs and improves in ease of use. Among the multiple features of Big Data, the use of nowcasting econometric techniques for economic scopes represents an important opportunity, which must be deeply analyzed by firms' managers. In fact, Big Data are considered by some scholars (Tene et al., 2013) the next management revolution for many economic sectors, and they are deemed pivotal for scopes as efficient social marketing, narrow segmentation, tailored products, and improved decision making by both researchers (McGuire et al., 2012) and firms (Russom, 2011). The technology still faces early-phase issues and debates concerning different areas. In fact, some scholars found possible harms for privacy (Buhl et al., 2013), which is still one of the most sensitive topics as confirmed by NSA recent scandals, and for data collection transparency (Richards et al., 2013), whose worsening could be caused by multilayered structures behind the collection process. Moreover, the "incremental effect" (Tene et al., 2013) of data gathering will lead to several issues with customers identity and profiling management, since it will be difficult to disentangle the two elements. Such controversial arguments require further analysis and research, as we still lack hard data for many hypothesis.

The study showed a new technique called nowcasting, which was conceived and mainly used for meteorological purposes (Mass, 2011) to predict thunderstorms, applied instead to social and business purposes. It is able to execute very short-range predictions, which differentiates it from common forecasting. Is it possible to apply such mechanics to

society through use of Big Data? The Predpol, compared with IBM Blue CRUSH's results, and the national mood prediction cases seem to confirm it. The former is a business software aimed at police departments, which allow them to receive predictions about hot areas in terms of crime activities. It resulted in double accuracy of predictions with respect to traditional crime analysts, and 13% reduction in crime rates in the first 4 months of adoption. The results are even more outstanding when compared with Blue CRUSH outcomes, which confirm that nowcasting applied to crime and societal issues does create benefits that are not severely contingent. In fact, Memphis PD enjoyed 30% crime reduction in 6 years of software adoption, avoiding extra costs and investments while optimizing existing resources. Unfortunately, the application of nowcasting still requires experiments and hard data in order to understand its magnitude of effect on other society sectors.

Finally, the last case shows the tool's flexibility, as a team of researchers were able to roughly predict the mood of the United Kingdom through the analysis of real-time online social messages or tweets (Lansdall-Welfare et al., 2012). The algorithm neatly linked riots to signs of national anger, while it was at the same time deceived and confirmed during holidays, when it noticed national joy, caused by wrong tweets' words interpretation.

The overall research highlights the need for further evaluation of nowcasting and Big Data applied to society and business environments, as the effects are still foggy and controversial. It is still possible that the hypothesis of important advantages for Big Data and nowcasting adopters will be nulled by the rise of severe harms to people's privacy and buying power, as the power of technologies' owner and user increases. This scenario calls for careful analysis of future outcomes.

References

- Banbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting.

- Boyd, D., & Crawford, K. (2011). Six provocations for big data.

- Buhl, H. U., Röglinger, M., Moser, D. K. F., & Heidemann, J. (2013). Big Data.*Wirtschaftsinformatik*, *55*(2), 63-68.

- Bureau of Justice Statistics (2003). Justice expenditure and employment in the United States, 2003. *US Department of Justice.*

- Castle, J., Hendry, D., & Kitov, O. I. (2013). *Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview* (No. 674).

- Chalfin, A. (2013). The economic cost of crime. *Encyclopedia of Crime and Punishment*

- Dixon, M., & Wiener, G. (1993). TITAN: Thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, *10*(6), 785-797.

- European Commission (2012). Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses. *Press release.*

- Gattiker, A. E., Gebara, F. H., Gheith, A., Hofstee, H. P., Jamsek, D. A., Li, J., ... & Wong, P. W. (2012). Understanding system and architecture for big data.*IBM research*.

- Press, G. (2014). Top ten Big Data pure-plays 2014. *Forbes* http://www.forbes.com/sites/gilpress/2014/02/11/top-10-big-data-pure-plays-2014/

- IBM (2011). Memphis PD: keeping ahead of criminals by finding the hot spots. http://www.ibm.com/smarterplanet/us/en/leadership/memphispd/

- Kelly, J. (2014). Big Data vendor revenue and market forecast 2013-2017. *Wikibon.* http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017

- Laney, D. (2011). 3D Data Management: Controlling Data Volume, Velocity and Variety. *Gartner*.

- Laney, D. (2012). The Importance of 'Big Data': A Definition. *Gartner*.

- Lansdall-Welfare, T., Lampos, V., & Cristianini, N. (2012). Nowcasting the mood of the nation. *Significance*, *9*(4), 26-28.

- Lütticke, M. (2013). A chronology of the NSA surveillance scandal. *DW* http://dw.de/p/1A9uu

- Mass, C., & Mass, C. F. (2011). Nowcasting: The Next Revolution in Weather Prediction. *Bulletin of the American Meteorological Society*.

- Mattmann, C. A. (2013). Computing: A vision for data science. *Nature*,*493*(7433), 473-475.

- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.

- McGuire, T., Manyika, J., & Chui, M. (2012). Why Big Data is the new competitive advantage. *Ivey Business Journal*, *7*(8).

- Met Office (2011). Nowcasting.

  http://www.metoffice.gov.uk/learning/science/hours-ahead/nowcasting

- Predpol Blog (2014). LAPD credits Predpol for no crime day. Predpol.

  http://www.predpol.com/lapd-division-credits-predpol-for-no-crime-

- Richards, N. M., & King, J. H. (2013). Three Paradoxes of Big Data. *Stanford Law Review Online*, *66*, 41.

- Rubinstein, I. S. (2012). Big Data: The End of Privacy or a New Beginning?.

- Russom, P. (2011). Big data analytics. *TDWI Best Practices Report.*

- Smith, C. (2014). How many people use 416 of the top social media apps?. *DMR*

  http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/#.U3iMh_I_skU

- Tene, O., & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics.

- www.predpol.com