UDK: 324:007.3
Pregledni rad
Primljeno 19. prosinca 2018.

Oraclum Intelligence Systems Case Study: Usage of Big Data in the Political Elections Prediction

MIRELA HOLY
Veleučilište VERN'
mirela.holy@vern.hr

Oraclum Intelligence Systems (OIS) is a company that uses data from social media and analyses this information toward prediction of election outcomes. The company has successfully predicted the outcomes of Brexit and the USA presidential elections. Precisely, the company analysts correctly predicted almost all swing states (Pennsylvania, Florida, North Carolina, and Ohio), and even predicted that Hillary Clinton could get more votes but would lose the Electoral College vote. This paper presents the case study of OIS successful and very accurate prediction of Brexit result as well as Trump's victory, answers how OIS used big data in research and why its research methods were successful. Considering the recent disclosure of Cambridge Analytica's big data misuse, the paper also focuses on the protection from similar misuses, particularly in the context of the General Data Protection Regulation (GDPR) entry into force.

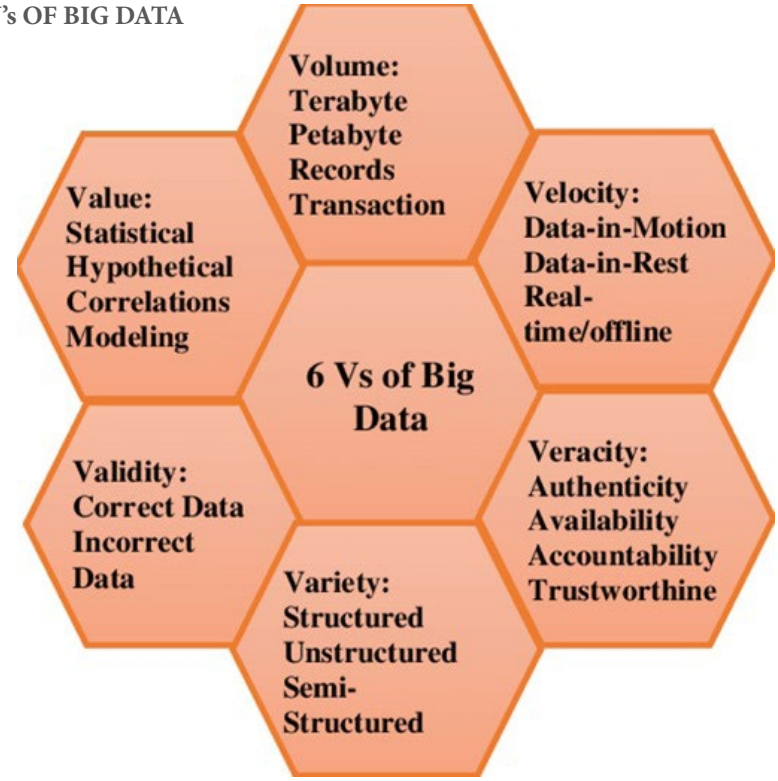
Keywords: big data, election prediction, Trump, Clinton, Brexit, GDPR

1. Introduction

Big data has become a very intriguing topic not only in information and computing sciences but also in social and political research. Moreover, some theorists write about big data as a phenomenon that will completely change the scientific paradigm. In the frame of Kuhn's idea of shifting scientific paradigms (1962/2013), big data is brought into the context of the so-called data revolution, i.e. proclamations of the fourth scientific paradigm (Hey et al., 2009; Kitchin, 2016).

But what is big data in fact? According to Snijders et al. (2012: 1) "big data is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software". Although the term big data came into use in the mid-1990s (Kitchin, 2016), which is quite a while ago, there is still no single definition accepted by most scholars. Big data is often associated with different acronyms that try to define big data by some characteristics. The term big data is often described by the following six characteristics: value, volume, velocity, variety, veracity and variability that are denoted as 6V's of big data (Ristovski and Chan, 2018).

Graph 1: 6V's OF BIG DATA



Source: Rahman et al., 2016: 4.

These 6V's in the acronym grew from the initial 3V's that referred to the following characteristics: huge in volume, high in velocity, and diverse in variety in type, structured and unstructured in nature (Laney, 2001; Kitchin, 2016). But today, practitioners and theorists use even larger numbers in the acronym to describe big data. For example, Firican (2017) writes about 10V's of big data in which he counts the following characteristics: volume, velocity, variety, variability, veracity,

validity, vulnerability, volatility, visualization and value. All acronyms connected with the big data nevertheless have three common characteristics: volume, velocity and variety so it can be concluded that something described as big data must include extremely large amount of data, must have exceptional speed at which data is being generated, and must implicate not only structured, but also semi-structured as well as unstructured data.

The key to understanding big data is data analytics that enables simplification of "the complexity of data and calculation for achieving an expected pattern of data sets and outcome" (Rahman et al., 2016: 4). For the usage of big data, consequently, it is very important to establish a logical framework for data analytics, which must consist of the following phases: initial planning, implementation and evaluation (Rahman et al., 2016). Given the fact that all projects, regardless of whether they are scientific or business, imply realization through these three logical phases, one may have questioned the above-mentioned thesis that big data represents the fourth phase of scientific revolutions. However, Anderson (2008) claims that big data signalled a new era of knowledge production characterized by "the end of theory" because "the data deluge makes the scientific method obsolete" (Kitchin, 2016). Anderson is not the only one who thinks that big data and data science will completely change science. Kitchin (2016) quotes a whole series of big data researchers who claim the same, for example, Prensky (2009) argues that "scientists no longer have to make educated guesses, construct hypotheses and models, and test them with data-based experiments and examples". Now, Prensky states, "they can mine the complete sets of data for patterns that reveal effects, producing scientific conclusions without further experimentation". Similarly, Dyche (2012) states that "mining big data reveals relationships and patterns that we didn't even know we should look for". Despite such claims, big data practitioners still use science as usual, or business as usual (research) project matrices in carrying out research.

Big data implies interdisciplinary research in application of different sciences and technologies such as mathematics, artificial intelligence, especially machine learning, data mining, cloud computing, real time data streaming technology etc. (Zhang, 2013; Rahman et al., 2016). This knowledge is used to create the most popular big data tools in the following areas: data storage and management, data cleaning, data mining, data analysis, data visualization, data integration, data languages and data collection (Rahman et al., 2016).

Table 1: MOST POPULAR BIG DATA TOOLS

No.	Big data Area	Tools
1.	Data Storage and Management	Hadoop, Cloudera, MongoDB, Talend
2.	Data Cleaning	OpenRefine, DataCleaner
3.	Data Mining	RapidMiner, Teradata, FramedData, Kaggle
4.	Data Analysis	Qubole, BigML, Statwing
5.	Data Visualization	Tableau, Silk, CartoDB, Chartio, Plot.ly
6.	Data Integration	Blockspring, Pentaho
7.	Data Languages	R, Python, RegEx, XPath
8.	Data Collection	Import.io

Source: Rahman et al., 2016: 7.

Big data nowadays is big business, which is noticeable in the number and variety of tools available. Also, apart from the firms that are creating big data tools, today a whole range of data analytic firms

are involved in providing big data research and make big profits on these projects. Keeping this in mind, the following Kitchin's (2016) conclusion seems true:

> There is little doubt that much of the rhetoric concerning big data is hyped and is boosterish, especially that produced by companies seeking to push new big data products, or research centres seeking to capture grant income. At the same time, there is no doubt that big data are qualitatively different to traditional small data and it does offer the potential to change how business is conducted, societies governed, and academic research carried out. Big data and new data analytics do offer the possibility of reframing the epistemology of science, social science and humanities (though it will not lead to the 'end of theory'), and such reframing is already actively taking place across disciplines (Kitchin, 2016).

Finally, it is important to emphasize that big data analysis has noticed constraints and inconsistencies. Zhang (2013) found four types of inconsistencies in big data: temporal inconsistencies, spatial inconsistencies, text inconsistencies and functional dependency inconsistencies and concluded that "articulating explicitly the types of inconsistent phenomena in big data can help pave the way to improve the quality of big data analysis". Big data is one of the fastest growing areas in the research field and the broad application of big data research suggests that big data will be an extremely important tool in various forms of marketing research, especially the political market research. But like many other research methods big data should also not be interpreted as a magic stick that will completely change marketing research field.

2. Usage of big data in political research

The development of the role and significance of big data research took place in parallel with the development and growing popularity of social media.[1] Precisely because of this connection, big data and social media are intensively explored in the context of big data research on huge amounts of data generated through social media. McCay-Peet and Quan-Haase concluded that

> Because of its proliferation in society as well as its unique technological affordances, social media provides new avenues for researchers across multiple disciplines, including health sciences, sociology, and political science, to collect rich, vast, and networked data, recruit diverse groups of participants and perform complex analyses (2016: 13).

Life nowadays is characterized in a social, business and political context by communication through social media, which have become the main media of information and communication for many people. The growth of the role and importance of social media has reduced the importance and social influence of traditional mass media in the last ten or fifteen years. This change of social and communication trends is also reflected in the sphere of politics. The previous focus of political, communicational and sociological research has shifted from the topic of mass media and politics, onto the influence that social media have on voters. McCay-Peet and Quan-Haase (2016: 13) emphasized that "data derived from user-generated content, such as posts, 'likes', and connections signalled through 'friends' and 'follows', have become central to many areas of study, including politics" (acc. Rainie et al., 2012).

1 "Social media are web-based services that allow individuals, communities, and organizations to collaborate, connect, interact, and build community by enabling them to create, co-create, modifies, share, and engage with user-generated content that is easily accessible." (McCay-Peet and Quan-Haase, 2016: 16)

The great potential of social media was recognized in terms of managing political communication, respectively the placement of political messages and ideas; discussing, persuading and motivating the electorate; discrediting political opponents; criticizing and ridiculing the ideas and programs of political competitors etc. In this way, social media have become a powerful instrument of political propaganda, to the same extent of the white, grey and black[2] political propaganda, which in the context of this media is recognizable under the term fake news.[3]

The usage of social media in a political context is not only through placement of information that aims to influence the attitudes and behaviour of voters and, consequently sharing and liking of this information and thus creating a perception of support for an idea, option, candidate, or program, but social media are also used to research the political preferences of social media users, and this area has great potential for big data researches. A special segment of big data research in the political context is devoted to prediction of election or referendum results through the analysis of social media communication. It is important to emphasize that "social data analytics is not only informing, but also transforming existing practices in politics" (Buus Lassen et al., 2016).

The influence of public opinion surveys results on the final election outcomes, and on the democratization of society in general, has been continuously present in political theory and practice since the first Gallup polls in the 1930's (Berinsky, 2006). Surveys of voter preference are not a new phenomenon, but with the trend of Americanization of politics, these types of research have had an increasing importance in political processes all over the world. Their popularity and significance is growing even in the USA so "the war surveys" marked the US presidential election in 2016 (Leetaru, 2016). Althaus (2003: 313) warns that the dynamics of preference aggregation presents two issues: (1) to what extent opinion surveys are seen as irrelevant to political decision making and (2) to what extent surveys are seen as the accurate measures of public opinion. With the first issue "we risk mistaking the public's reasoned judgment with the result that democratic project must surely suffer. With the second one "we risk misreading the voice of the few as the will of many". Despite the constraints of public opinion polls, in spite of being publicly known that public opinion polls could be used as a means of political propaganda, the fact is that media and the public uncritically consume this information that can still have a small impact on the process of political decision-making (Lamza Posavec and Rihtar, 2007).

Although traditional political preferences surveys conducted through field or telephone polling methods can also manipulate survey results in favour of a certain political option or candidate,[4] political preferences surveys through social media lead to a higher level of suspicions regarding political favouritism because of the possibilities that provide abuse of big data research. Although the ethical aspects of using and analysing big data generated on social media have been debated for a long time, especially in the context of user privacy violations, the Cambridge Analytica case has opened new ethical questions regarding big data manipulation on social media. Cambridge Analytica was a data analytics company that worked with Donald Trump's election team and the winning Brexit campaign. Company used millions of Facebook profiles of US voters to build a powerful software

2 According to Jowett and O'Donnell (2012: 17-19) white propaganda comes from a source that is identified correctly, and the information in the message tends to be accurate. Black propaganda is when the source is concealed or credited to a false authority and spreads lies, fabrications, and deceptions. Black propaganda is the "big lie," including all types of creative deceit. Grey propaganda is somewhere between white and black propaganda. The source may or may not be correctly identified, and the accuracy of the information is uncertain.

3 According to Cambridge dictionary fake news are "false stories that appear to be news, spread on the Internet or using other media, usually created to influence political views or as a joke".

4 For example, when selecting a sample within one constituency, it is possible to choose the areas where the political option for which the survey results are adjusted has a better result that creates a false image (Leetaru, 2016).

program to predict and influence choices at the ballot box. Christopher Wylie, Cambridge Analytica whistle-blower, revealed that data was collected through an app called *thisisyourdigitallife*. This app was developed by academic Aleksandr Kogan who, through his company Global Science Research (GSR) in collaboration with Cambridge Analytica, paid hundreds of thousands of users to take a personality test and agree to have their data collected for academic use. But their data, as well as personal data of their Facebook friends, was not collected for academic, but for political purposes to manipulate US voters in 2016 presidential elections (Cadwalladr and Graham-Harrison, 2018). Cambridge Analytica's procedures for its clients thus went far beyond the framework of standard research of voter preference or political community management and represent a serious threat not only to the privacy of users but also of the even softest limits of ethical political communication.

Given the fact that potential misuse of personal information generated not only by social media, but even to a far greater extent by search engines such as the IT giant Google, as well as other private or public organisations, has become a subject of great public concern, the European Union has decided to legislate this matter. At the end of May 2018, the General Data Protection Regulation (GDPR) came into force in all EU Member States. GDPR is an extremely important regulation that would thoroughly change the way of using personal data by economic and political entities. Protecting personal data is one of the basic tasks that GDPR puts in front of organizations whether it is personal data of customers, clients or employees. Organizations must know at all times where these data are stored and for what purpose they may be used. Likewise, in the event that someone decides to withdraw the privilege of using his or her personal data, the organization must be able to do so within the set deadline. Failure to comply with the GDPR provisions entails very serious penalties - up to 4% of total annual income worldwide or up to 20 million euros, whichever is higher. GDPR applies to all businesses operating in the EU - including microenterprises, small and medium-sized enterprises, public institutions, bodies and agencies that collect personal data. According to GDPR personal date are the following: name, address, e-mail address, IP and MAC address, GPS location, RFID tags and cookies on websites, phone number, photos, video footage of individuals, OIB, biometric data (fingerprint, eye data), genetic data, educational and professional information, pay data, credit information, bank account details, health data, sexual orientation, voice, and many other data pertaining to an individual whose identity has been identified or can be determined.[5]

Oraclum Intelligence Systems is a private firm registered in Cambridge, UK, to provide services such as forecasting political results, predicting product demand, real time big data analysis in retail, microtargeting customers, optimizing the marketing mix, and experiments and validated learning. Oraclum Intelligence Systems predicted very precisely the results of the US presidential election in 2016 and the results of Brexit, the British referendum regarding EU membership, also in 2016. At the first glance, Oraclum Inteligence Systems (below the text OIS) may seem similar to Cambridge Analytica (CA) because OIS provides similar services to those of the notorious CA. Given the fact that after the Cambridge Analytica scandal a negative image of data analytic companies prevails, this case study of OIS was motivated by the question whether it is possible to use big data political analysis in a morally acceptable way.[6]

5 https://gdpr2018.eu/sto-donosi-gdpr/, 24.8.2018.
6 http://www.oraclum.co.uk/home/, 24.8.2018.

3. Research framework

The problem of this study is insufficient research of the big data methods of the company Oraclum Intelligence Systems in the field of political polls, especially the reasons why these methods are more accurate than others.

The subject is a case study of Oraclum Intelligence Systems, the company that predicted the results of the US presidential election and Brexit, both in 2016.

Research objectives and questions were the following:

RQ1: to determine exactly what method of research was used by the OIS;

RQ2: to determine why the OIS research methods were more precise than other research of US and British voter preferences during 2016 elections / referendum;

RQ3: to determine the role of big data in these pieces of research;

RQ4: to determine the future of such research as well as its ethical and legal constraints.

The research was conducted using the following methods:

(1) secondary data analysis, i.e. an analysis of relevant literature regarding big data and political market research;

(2) an analysis of OIS web site content and media coverage of OIS;

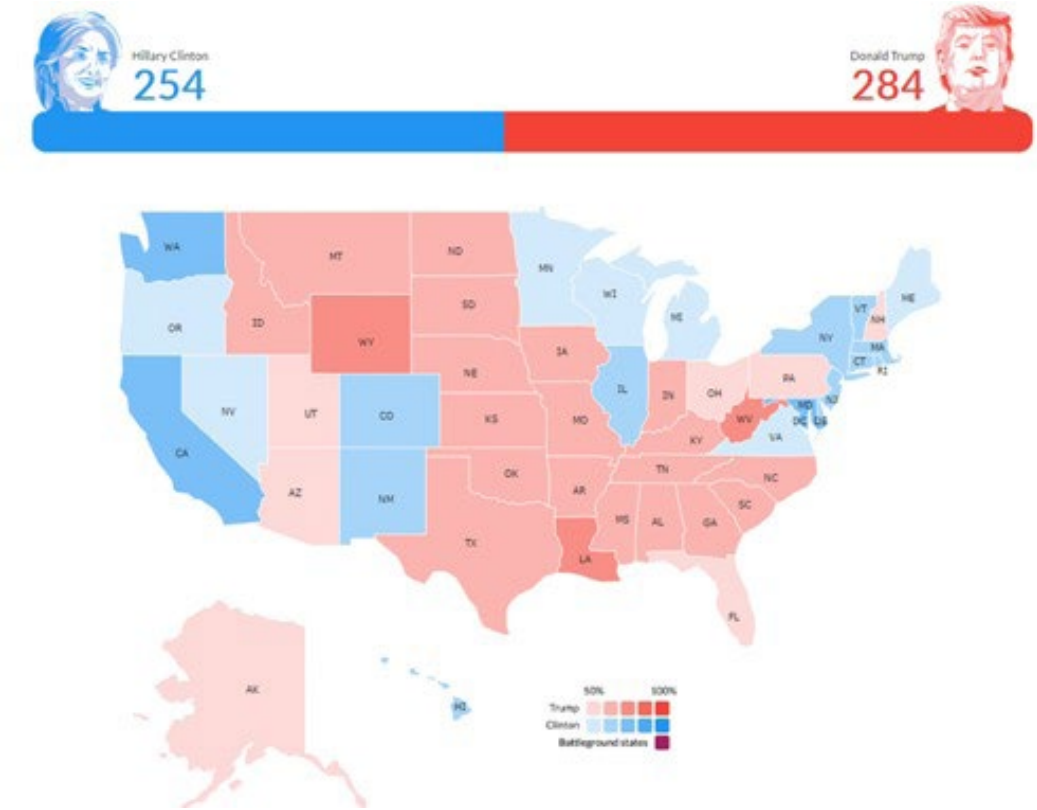
(3) an in-depth interview with Vuk Vuković, co-founder of OIS (eight groups of questions on March 26, 2018).

4. About Oraclum Intelligence Systems (OIS)

Oraclum Intelligence Systems (OIS) is a company that uses data from social media and analyses this information with the help of big data software toward prediction of election outcomes and discovering patterns of consumer behaviour. The company's election data analysis has successfully predicted the outcomes of Brexit and the USA presidential elections. Precisely, the company analysts correctly predicted results in swing states such as Pennsylvania, Florida, North Carolina, and Ohio, and even predicted that Hillary Clinton could get more votes but, finally, lose the Electoral College vote. The success of OIS was a surprise because they have accurately predicted the winners in all of swing states except two (Michigan and New Hampshire). For example, Pennsylvania, for which not a single survey gave advantage to Trump, OIS predicted with incredible precision. They gave Trump 48.2%, and Clinton 46.7%. The real score was 49.1% for Trump against 47.7% for Clinton. Similar precision was demonstrated in case of Florida and North Carolina, for which some researchers did not dare to make predictions because they were "too close to call".[7]

7 http://www.oraclum.co.uk/about/we_really_did_predict_trump/, 21.8.2018.

Graph 2: OIS PREDICTION SURVEY OF US 2016 PRESIDENTIAL ELECTIONS



Source: http://www.oraclum.co.uk/about/we_really_did_predict_trump/, 21.8.2018.

OIS achieved such precise prediction of US presidential election results based on a very small sample of 445 respondents. This was possible owing to the research methodology applied by OIS, which is not reduced to the usual quantitative methods, but to a combination of several methods that combine quantitative and qualitative research.

Equally impressive was their prediction of the results of the Brexit referendum. Their research showed that the result would be 51.3% for leave, and the actual outcome was 51.9%.[8] Their predictions had been published by several media in Croatia, UK and USA before the final results, which can be checked by looking into the contents of these articles.[9] For example, New Scientist published the article "Time is ripe for a new science of polling"[10] on November 3rd, six days before the elections. Jutarnji list published the article "Respectable Croatian Economist: The new American president will still be Donald Trump"[11] on November 8th, one day before the elections. It was similar with Index.hr who published the article "Croatian economist whose company predicts election results: "Trump will

8 http://www.oraclum.co.uk/about/we_really_did_predict_trump/, 21.8.2018.
9 http://www.oraclum.co.uk/about/media/, 21.8.2018.
10 https://www.newscientist.com/article/2111298-as-us-election-looms-time-is-ripe-for-a-new-science-of-polling/, 21.8.2018.
11 https://www.jutarnji.hr/vijesti/americki-izbori/ugledni-hrvatski-ekonomist-novi-americki-predsjednik-ipak-ce-biti-donald-trump/5239047/, 21.8.2018.

win, here's how"[12] on November 8th, day before elections. Vuković adds that OIS, apart from these predictions, also successfully predicted the French presidential election, or better said, the absolute victory of Macron, as well as the results of the first round of the 2017 Zagreb mayor elections. OIS predicted a month before the elections 31.2% for Bandić (versus 30.8%), 25.1% for Mrak-Taritaš (versus 24.5%), and 17.9% for Švaljek (versus 19.1%). At the same time two Croatian survey agencies favoured Mrak Taritaš over Bandić and stated the uncertainty as to whether Švaljek or Bandić would enter the second round.

But along with these successful predictions, OIS also had a miss – the prediction of repeated Croatian elections in 2016, when OIS predicted the victory of the SDP coalition over HDZ, but just the opposite happened (instead of 61:54 for the SDP coalition it was 61:54 for HDZ). Then OIS team realized that the pattern was too biased because they had too many SDP voters in the poll and too few HDZ voters, which even their online analysis could not correct (because the whole group bias leaned too much towards SDP).

The company was founded by PhD Mile Šikić, PhD Dejan Vinković, and Vuk Vuković. After receiving his PhD in physics from the University of Kentucky, Vinković moved to the Institute for Advanced Study in Princeton. He became a professor at the Physics Department at the University of Split, Croatia. He also accepted the position of Chief Science Officer at Hipersfera Ltd. Šikić is an associate professor at the Department of Electronic Systems and Information Processing at Zagreb University, the Faculty of Electrical Engineering and Computing, and a visiting scientist at the Bioinformatics Institute, A*STAR Singapore. Vuković is a PhD candidate at the University of Oxford, Pembroke College, where he also teaches Causal Inference at the Oxford Q-Step Centre. He holds a Master of Science degree from the London School of Economics, in the field of political economy.[13] Šikić, Vinković and Vuković got together at a conference organized by Vinković in 2014 and found out that they share the same interests, such as monitoring of election results and analysing the correlation of certain socio-economic indicators and how people vote. They decided to do a project for the Croatian newspaper Jutarnji list and in a series of 16 articles over a period of one and a half months before the parliamentary elections 2015 they analysed the elections. They also experimentally tested their model of prediction and network analysis, and in mid-2016 they decided to launch their company.

According to the company's website OIS partners are UX Passion for app design, Cantab Analytica for consulting and Zenlab for data mining.[14]

12 https://www.index.hr/vijesti/clanak/poznati-hrvatski-ekonomist-cija-tvrtka-prognozira-rezultate-izbora-pobijedit-ce-donald-trump/930796.aspx, 21.8.2018.

13 http://www.oraclum.co.uk/about/team/, 21.8.2018.

14 http://www.oraclum.co.uk/about/team/, 21.8.2018.

5. Research results

RQ1: Oraclum Intelligence Systems Methods

According to Vuk Vuković, OIS uses the Bayesian Adjusted Social Network (BASON) Survey[15] that successfully solves the biggest issue in polling – it eliminates bias from respondents' answers. OIS analyses the social networks of customers (or voters) in order to draw inferences from these networks and predict outcomes with great precision. OIS BASON survey is the only poll in the world that successfully predicted both Brexit and the victory of Donald Trump. BASON combines the wisdom of crowds (WoC) approach to polling adjusted for the network analysis of groupthink bias with the idea to incorporate wider influences, including peer groups that shape an individual's choice on voting day. BASON doesn't require a representative sample and goes beyond representativeness, self-selection problems and random sampling, and focuses simply on trying to find out how people estimate their local conditions and sentiments. The fundamental difference between OIS BASON method and other research methods is in questions that OIS posed to respondents and, later, in the analysis of the user's network. OIS asks three key questions:

- the first one is for whom the respondent will vote;
- the second one is what respondents think how much their preferred choice will get (in percentages) in their constituency;
- the third one, their opinion on how people in their neighbourhood estimate that other people around them will vote.

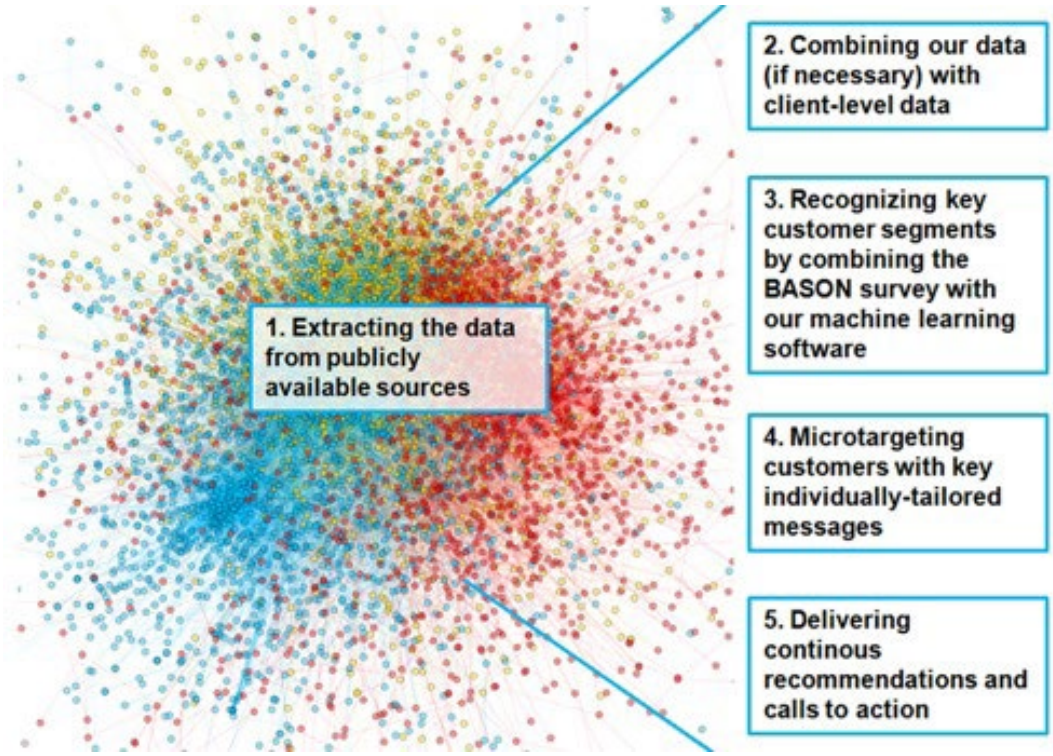
These questions are intentionally complicated in order to make the respondents think, and for this reason OIS consciously sacrifices the number of respondents to precision. They, finally, have relatively small samples because people give up. Their method then mathematically defines the most likely percentage of votes for a particular candidate. They also use mathematical estimates of how subjective the participant was in the assessment, i.e. whether the answer was realistic. This requires a network analysis. It allows OIS to measure the bias of each respondent. This part of survey - uncovering bubbles – has the goal to reduce the individual-level bias (confirmation bias) of each survey respondent through the network analysis. OIS looks for clustering within groups based on expressed preferences, which enables recognition of echo chambers, members of which are less likely to be good forecasters.[16]

The wisdom of crowds is the first part of OIS survey, and it is contained through the above-mentioned three key issues. It is necessary to get information about people's opinion on who will win. Namely, a person can support one party but can also see that the environment thinks differently and that person's party or candidate will not necessarily win. OIS survey is an aggregator of everything that voters know about choices - from the media, from their friends and acquaintances, from knowing the history of voting in their communities, etc. The wisdom of crowds represents a significant improvement over regular surveys both in accuracy and information gained.

The data science and analytics part is the key final component of both of OIS main products, BASON and the Wisdom of crowds. OIS combines BASON with data science through the following five phases:

15 http://www.oraclum.co.uk/products/, 21.8.2018.
16 http://www.oraclum.co.uk/products/bason/, 24.8.2018.

Graph 3: THE OIS APPROACH (COMBINING BASON WITH DATA SCIENCE



Source: http://www.oraclum.co.uk/products/data-science-and-analytics/, 24.8.2018.

However, Vuković explains, that their services of data science and analysis can also be used separately, for clients who handle large quantities of data on a daily basis and in real time. OIS can help clients in the structure of their data, combine it with OIS powerful tools, and tell the client how to get the most out of it to grow their business.[17]

RQ2: Why OIS research methods were more precise than others

Vuković explains that network analysis is the main reason why their method works. The sole usage of the wisdom of crowds method or more precisely those three questions, would not be enough to accurately predict neither the results of Brexit nor the presidential elections result.

The OIS idea is the following: anyone who can fill out their online survey can share it through social media (Facebook, Twitter) and thus attract their friends to the survey. If friends fill out the poll, the OIS team can see how clusters are created within the network and how people are connected within the network. In this way, OIS can identify those respondents who usually associate with homogeneous group that are rarely exposed to different attitudes from their own (this is called an echo chamber or simply a bubble). Those who are part of such an echo chamber will be less precise in their predictions than those who are in heterogeneous groups. Based on that, OIS has developed the measurement method for each respondent where OIS gives greater importance to those whom the network recognizes as more precise. To the contrary, OIS gives less importance to those who are a part of the echo chamber.

17 http://www.oraclum.co.uk/products/data-science-and-analytics/, 24.8.2018.

Vuković claims that OIS absolutely respects user privacy. OIS doesn't know neither it attempts to find out who has voted for which candidate or party (OIS gets an anonymous ID from each survey user). OIS's business model is not based on someone's personal information, but solely on what people say - in their opinion, and in the opinion of their friends.

RQ3: The role of big data in OIS political research

Vuković explains that big data is not so important for OIS in political research, but rather as a tool for business users, especially with the unstructured data that a client gives OIS. A client provides OIS with accesses to their own unstructured database from which OIS then draws out concrete patterns of behaviour and analyses it.

For example, retailers can track millions of customer accounts and thus predict which products sell better when put together, and find out the underlying trends. It is not possible to know who bought the product - this is only possible with loyalty cards - but without a loyalty card it is possible to draw very precise predictions about when customers will buy something and why. Such things are very useful as real-time analyses as they can achieve significant savings. Vuković concludes that with big data analysis it is possible to plan better.

RQ4: Usage of big data in the future

According to Vuković as a practitioner, but also as a scientist and theorist, big data will have a significant impact in the future, and not just big data, but all other technological tools that are growing based on the availability of data. Vuković especially emphasizes the Internet of things and the 5G Internet that will allow everyone in their own home to be connected to the Internet, creating every second a crowd data, to the potentially revolutionary blockchain technology that could completely eliminate any intermediaries between buyers and sellers, and significantly decentralize decision making (and thereby increase consumer power), all with a high level of transparency and security.

Considering these trends, he thinks, it is important that students, who are learning about big data, also learn how to do programming, and how to use statistics. These competences are crucial for a high-quality analysis of large amounts of data. Companies will have a growing need for the expertise of a data scientist, concluded Vuković.

6. Conclusion - ethical and legal constraints of big data political research

Although big data is very an attractive topic for interdisciplinary research, big data methods in political market research are still insufficiently explored area. The ethical and legal aspects of the use of personal data in big data research are getting better explored but still insufficiently. As recent international communications market research showed that the usage of big data as a PR tool in the future would constantly increase,[18] there is a great need for further exploration of the application of big data within political and communicational research, particularly ethical issues that arise from such research.

This case study of the Oraclum Intelligence Systems, company that precisely predicted the results of the US presidential election and Brexit, showed that OIS combines qualitative and quantitative research methods in predicting results of political elections. More precisely, just a combination of several research methods and disciplines (mathematics, artificial intelligence - machine learning and data mining) is the reason why OIS was far more successful and precise in predicting political outcomes then other agencies and data analysis companies. Although Vuk Vuković, the co-founder of the company, claims that big data does not have a big role in their political research, the fact is that he also claims that the network analysis is the main reason why their method is so accurate. Without applying big data tools, it would be almost impossible to analyse the echo chambers or bubbles on social media, because big data tools allow the recognition of echo chambers, or respondents who are less likely to be good forecasters. Consequently, it can be concluded that big data has a major role to play in the precision of research of political preferences conducted by the OIS. Further, there is no doubt that big data analysis will have an important role in the future political surveys. Moreover, it can be concluded that in the future it will be almost impossible to implement complex political research without the big data analysis. But big data primarily need to be seen just as another useful method, not as a new phase in scientific revolutions. While it is true that applying big data in research can lead to totally unexpected results, it is equally true that applying "traditional" research methods did not once result in revolutionary, totally unexpected cognitions. The logical structure of all research, regardless of the method, will remain unchanged. Finally, the recent case of Cambridge Analytica has shown that the unscrupulous usage of the social media users personal data, analysed with the help of modern technologies, in particular big data, is contrary to the public interest. Because of this it is good that GDPR has come into force. But, big data is a very powerful political marketing tool that in itself is neither good nor bad. In other words, big data can be used for useful but also for manipulative purposes. It is therefore extremely important that those who use big data in research are aware of their own moral responsibility during the application of this instrument, that they do not cross the line that endangers the privacy of data owners and social media users and that they do not use this instrument to manipulate voters.

18 https://annenberg.usc.edu/research/center-public-relations/news/usc-annenbergs-2017-global-communications-report-predicts, 24.8.2018.

Bibliography

Althaus, S.L. (2003): *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*, Cambridge: Cambridge University Press.

Anderson, C. (2008): *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, https://www.wired.com/2008/06/pb-theory/ (21st August 2018)

Berinsky, A.J. (2006): American Public Opinion in the 1930s and 1940s: The Analysis of Quota-Controlled Sample Survey Data, *Public Opinion Quarterly*, 70 (4): 499–529

Buus Lassen, N., la Cour, L. and Vatrapu, R. (2016): Predictive Analytics with Social Media Data. In Luke Sloan and Anabel Quan-Haase (eds.), *The Sage Handbook of Social Media Research Methods*: 328-341, Thousand Oaks, CA: SAGE Publications Ltd.

Cadwalladr, C. and Graham-Harrison, E. (2018): *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (22nd August 2018)

Dyche, J. (2012): *Big Data 'Eurekas!' Don't Just Happen*, https://hbr.org/2012/11/eureka-doesnt-just-happen, (21st August 2018)

Firican, G. (2017): *The 10 Vs of Big Data*, https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx, (21st August 2018)

Jowett, G.S. and O'Donnell, V. (2012): *Propaganda and Persuasion, Fifth Edition*, Thousand Oaks, CA: SAGE Publications, Inc.

Kitchin, R. (2016): Big Data – Hype or Revolution? In Luke Sloan and Anabel Quan-Haase (eds.), *The Sage Handbook of Social Media Research Methods*: 27-38, Thousand Oaks, CA: SAGE Publications Ltd.

Kuhn, T. (2013): *Struktura znanstvenih revolucija*, Zagreb: Naklada Jesenski i Turk / Hrvatsko sociološko društvo

Lamza Posavec, V. and Rihtar, S. (2007): Mogući utjecaj objavljenih rezultata predizbornih anketa na izborno ponašanje birača, *Društvena istraživanja* 16 (1-2): 87-88

Laney, D. (2001): *3D Data Management: Controlling Data Volume, Velocity and Variety*, https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf, (21st August 2018)

Leetaru, K. (2016): *What Opinion Polls Teach Us About Data Versus Truth*, https://www.forbes.com/sites/kalevleetaru/2016/11/08/what-opinion-polls-teach-us-about-data-versus-truth/#41c87e56749f, (22nd August 2018)

McCay-Peet, L. and Quan-Haase, A. (2016): What is Social Media and What Questions Can Social Media Research Help Us Answer? In Luke Sloan and Anabel Quan-Haase (eds.), *The Sage Handbook of Social Media Research Methods*: 13-26. Thousand Oaks, CA: SAGE Publications Ltd.

Novak, T. (2016): *Ugledni hrvatski ekonomist 'Novi američki predsjednik ipak će biti Donald Trump'*, https://www.jutarnji.hr/vijesti/americki-izbori/ugledni-hrvatski-ekonomist-novi-americki-predsjednik-ipak-ce-biti-donald-trump/5239047/, (21st August 2018)

Prensky, M. (2009): *H. sapiens digital: From digital immigrants and digital natives to digital wisdom*, https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1020&context=innovate/, (21st August 2018)

Rahman, H. and Begum, S. and Ahmed, M.U. (2016): *Ins and Outs of Big Data: A Review. The 3rd*

EAI International Conference on IoT Technologies for HealthCare, https://www.researchgate.net/publication/308302879, (21st August 2018)

Rainie, L., Smith, A., Lehman Schlozman, K., Brady, H. and Verba, S. (2012): *Social media and political engagement*, http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_SocialMediaAndPoliticalEngagement_PDF.pdf, (21st August 2018)

Ristevski, B. and Chen, M. (2018): Big Data Analytics in Medicine and Healthcare, *Journal of Integrative Bioinformatics*, May 2018.

Snijders, C. and Matzat, U. and Reips, U.-D. (2012): "Big Data": Big Gaps of Knowledge in the Field of Internet Science, *International Journal of Internet Science,* 7 (1): 1–5

Zhang, D. (2013): *Inconsistencies in big data. Cognitive Informatics & Cognitive Computing (ICCI*CC)*, 2013 12th IEEE International Conference: 61-67, https://www.researchgate.net/publication/261448807, (21st August 2018)

Sources

Meaning of "fake news" in the English Dictionary (2018): *Cambridge Advanced Learner's Dictionary & Thesaurus © Cambridge University Press*, https://dictionary.cambridge.org/dictionary/english/fake-news, 21.8.2018.

Oraclum Intelligence Systems (2018): *Oraclum Intelligence Systems*, http://www.oraclum.co.uk/home/, (21st -24th August 2018)

R.I. (2016): *Hrvatski ekonomist čija tvrtka prognozira rezultate izbora: "Pobijedit će Trump, evo kako"*, https://www.index.hr/vijesti/clanak/poznati-hrvatski-ekonomist-cija-tvrtka-prognozira-rezultate-izbora-pobijedit-ce-donald-trump/930796.aspx, (21st August 2018)

Što donosi GDPR? (2018): Europska unija, https://gdpr2018.eu/sto-donosi-gdpr/, (24th August 2018)

Time is ripe for a new science of polling. (2018): *Newscientist.com*, https://www.newscientist.com/article/2111298-as-us-election-looms-time-is-ripe-for-a-new-science-of-polling/, (21st August 2018)

USC Annenberg's 2017 Global Communications Report predicts convergence of marketing and public relations (2017): *Annenberg School for Communication and Journalism*, https://annenberg.usc.edu/research/center-public-relations/news/usc-annenbergs-2017-global-communications-report-predicts, (24th August 2018)

Studija slučaja tvrtke Oraclum Intelligence Systems: Upotreba velikih podataka u predviđanju političkih izbora

MIRELA HOLY
Veleučilište VERN'
mirela.holy@vern.hr

Oraclum Intelligence Systems (OIS) tvrtka je koja koristi podatke s društvenih medija i analizira te informacije u svrhu predviđanja izbornih rezultata. Tvrtka je uspješno prognozirala ishode Brexita i predsjedničkih izbora u SAD-u. Upravo su analitičari ove tvrtke ispravno predvidjeli rezultate u gotovo svim oscilirajućim saveznim državama (Pensilvanija, Florida, Sjeverna Karolina i Ohio), pa čak i predvidjeli da bi Hillary Clinton mogla osvojiti više glasova, ali izgubiti izborne elektore. U članku je predstavljena studija slučaja njihovog uspješnog i vrlo točnog predviđanja rezultata Brexita i Trumpove pobjede, te se u članku objašnjava kako je tvrtka koristila velike podatke u istraživanjima i zašto su njihove metode istraživanja bile uspješne. Uzimajući u obzir otkrivanje zlouporabe podataka u aferi Cambridge Analytica, članak se također fokusira na zaštitu od sličnih zlouporaba, posebno u kontekstu stupanja na snagu Opće uredbe o zaštiti podataka (GDPR).

Ključne riječi: veliki podaci, predviđanje izbora, Trump, Clinton, Brexit, GDPR