

A Review of Selected Aspects of Big Data Usage in Banks' Risk Management

Ivana Dvorski Lacković

idvorski@foi.unizg.hr

*Faculty of Organization and Informatics
University of Zagreb, Varaždin, Croatia*

Vladimir Kovšca

vkovsca@foi.unizg.hr

*Faculty of Organization and Informatics
University of Zagreb, Varaždin, Croatia*

Zrinka Lacković Vincek

zslackovi@foi.unizg.hr

*Faculty of Organization and Informatics
University of Zagreb, Varaždin, Croatia*

Abstract

Contemporary banks operate in a changing environment influenced by regulatory requirements, emerging new risk types and high competition on the market. At the same time banks have on their disposal large datasets from internal and external sources – the question arising is whether banks are using adequate analytics to gain value from big data. Big data has been used in banking for some time, mostly in the marketing field, but the usage of big data in banks' risk management has not been a subject of extensive scientific research. The goals of this paper are twofold. Firstly, the authors are presenting an overview of existing research on the topic of big data usage in banks' risk management by discussing selected aspects of this topic. Secondly, the authors are formulating a proposal on how to strategically use big data in risk management in banks by merging literature based extracted usage of big data with risk management phases. Expected contributions of this paper are systematization of knowledge from available literature and increasing scientific and professional awareness on the topic and potential of big data in banks' risk management.

Keywords: big data, banks, risk management

1. Introduction

Big data technology has been an important topic in the information technology research area and has been used in banking for some time, mainly in marketing domain and as a tool for fraud detection but its potential in banking regarding risk management has only recently been discovered. Consulting companies that operate in financial domain are doing extensive research and disseminating knowledge about importance and potential of big data in risk management. On the other hand, scientific

literature on the subject is rather scarce. The topic of this paper is focused on review of some selected aspects of big data technology in banks' risk management due to several reason. Firstly, as we will point out in continuation of this paper, risk management is currently one of the most dynamic banking areas. Secondly, it presents not only a field in which regulatory requirements need to be satisfied, but also an opportunity for banks' management to make decisions based on banks specific individual risk profile and enhance their business activities.

The paper is divided into several sections. In first section the authors are presenting an overview of regulatory framework and current status of risk management in banks. Following is the section related to big data and it explores big data basics, drivers and impediments to big data adoption in banks. Third section is related to an overview of existing studies of big data and its usage in banks' risk management. Fourth section is related to exploration on how big data can be strategically used in banks' risk management. Finally, conclusions of research and guidelines for future research are presented.

2. Overview of regulatory framework and risk management practices in banks

Banks are nowadays faced with multiple business challenges. The role of a bank as an institution empowered to take clients' deposits and transform them into loans and make profit based on these actions is not a true replica of a contemporary bank and its tasks. Oracle research [1] suggests that "financial institutions are nowadays transient nodes for the customer who is the center of attention, in sense that he chooses his business relationships with a number of institutions that offer financial services depending on what he assesses as the most profitable combination in terms of what he gets from that relationship. Consumers also expect to have full transparency about the products and services being offered and thus for banks and financial services companies to keep customers for the long term, they must get closer to them and anticipate customer needs and be able to proactively position their products." Parallel, the rise in popularity of other types of financial institutions that offered quick and higher returns to clients' investment, made a shift in banks from traditional banking activities to diversification and engagement in riskier business activities. Combined with the challenge to offer new and innovative products and services to clients and to stay efficient at the same time, this led to increase in banks' risk exposures. Šverko [2, pp. 13] argues that "banks have actually become "producers" of risk due to fact that they are underwriting risks, transforming them and incorporating them into their products and services."

At the same time the evolution of information and communication technology led to significant changes and brought banks the ability to offer clients innovative and improved services. On the other hand, non-compliance with postulates of technology usage and the misuse of technology became a significant source of internal and external risks.

The cases of banks' large losses and later on the occurrence of global financial crisis led to the need to regulate the banking sector more strictly. This was achieved

through enhancement of existing regulatory standards in the domain of risk management for banks. In the European Union current legislative that is obligatory, all-encompassing and that regulates the process of risk management in banks is *The Regulation EU No 575/2013 of The European Parliament and of The Council on prudential requirements for credit institutions and investment firms and amending Regulation EU No 648/2012*, colloquially called Basel III, and its underlying acts. National supervisory banking bodies have harmonized its legislation, requirements imposed on banks operating on the national territory and its supervisory activities with the Basel III provisions. The scope of Basel III is related to:

- bank risk related definitions,
- capital requirements for credit, operational and market risk,
- limitation of large exposures,
- liquidity risk and interest rate risk management,
- financial leverage management,
- risk related reporting.

According to Miloš Sprčić [3] generally there are many definitions of risk in scientific and professional literature, most of which are focused on negative aspects of risk, i.e. potential losses organization is faced with due to certain events expressed in downside effects on expected earnings, cash flow or company value. But as [3, pp. 17] argues risk management can be approached more widely and encompass both positive and negative aspects of risk occurrence. In this paper we will follow this holistic approach to risk management and try to explore how big data can enhance risk management and have a positive effect on bank in a whole.

Banks are faced with various types of risks, depending on their scope of business activities. Banks that are offering innovative financial services or the ones using advanced financial and hedging techniques, are exposed to higher risks and require more sophisticated risk management methods and techniques. The simplest and basic classification of risks banks are faced with encompasses:

- *credit risk* – risk of losses due to borrowers default or rating worsening,
- *operational risk* – risk of losses due to internal or external events resulting from internal processes, people, systems or external events,
- *market risk* – risk of losses due to changes in instruments' price or exchange rate,
- *liquidity risk* – risk of losses due to inability to sell instrument quick enough on the market (market liquidity) or due to the fact that the company cannot pay its liabilities on time (funding liquidity),
- *interest rate risk in banking book* – losses due to effects of changed interest rate on banks' economic capital.

In order to achieve this goal and according to Basel III requirements, banks are obliged to organize three control functions within organization (i.e. three lines of defense):

- *risk management control* – this function is entitled to control how banks' units are managing risks they are exposed to, conduct stress testing, control limit breaches and deliver reports to internal and external users;

- *compliance* – the role of this control function is to ensure that the bank is conducting its activities in accordance with external regulations and internal procedures;
- *internal audit* – whose task is review and all-encompassing control of business segments, activities and processes in a bank.

This shortly presented overview of movements within banking sector and presentation of basic risk types banks are faced with, gives us an insight on the fact how wide the “playground” for risk management is and leads to the conclusion that special attention needs to be payed to adequate organizational hierarchy, establishment and documentation of internal procedures, development and promotion of organizational risk culture, development and control of risk limits and formulation of business continuity plans in order for risks to be managed adequately.

3. Big data

According to [4] big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. Originally big data was a term used to describe large datasets that could not be captured, stored, managed nor analyzed using traditional databases. With years the term broadened and now encompasses also the set of technologies that perform all the mentioned functions, varied collections that solve complex problems and make value from that data more economical [5].

IBM [6] defines the ‘four V’s’ behind big data mechanism:

- *volume* (quantity of data. i.e. massive data sets),
- *velocity* (accelerating speed at which data is being generated today),
- *variety* (increasing diversity in the types and sources of data requiring management and analysis),
- *veracity* (reliability of particular data).

Özköse et al. [7] argue that there are ‘five V’s’ behind the big data term, adding ‘*value*’ defined as the result generated from all the procedures and enriching the process to the before mentioned ‘four V’s’.

3.1. Drivers of big data adoption in banks

Mohanty et al. [8, pp. 54] argue that big data are applicable on several levels in the banking industry, including retail (branch and mortgage banking, credit card and private banking), commercial (credit risk analysis, non-credit products, client management and sales, middle market lending), capital markets (trading and sales, underwriting and structured finance, non-depository credit institutions) and asset management (wealth management, asset investment management, asset issuer services, global asset reporting and investment deposit analytics).

According to [5] drivers of big data technology adoption in the financial industry are:

- *explosive data growth* (increased number of banking instruments and transactions),
- *regulation* (requirements regarding the real-time view on risk management and related financial transactions; enhanced risk reporting; risk simulations),
- *fraud detection and security* (operational risk management issues, such as spotting fraud risk among its own employees or identifying rogue traders),
- *customer insight and marketing analytics* (understanding consumer behavior and interests and consequently shaping new products in accordance with customer needs).

According to McKinsey's research on big data and credit risk [9] there are six key trends that are expected to change bank risk management. These are:

- expanded regulatory requirements,
- use of technology and mathematics in risk management,
- changing customer expectations,
- de-biasing judgmental decision-making,
- emerging risk types and
- growing cost discipline.

When discussing risk management, regulatory changes are the key factor that has influence on business practice and are starting point for changes in business process. Regulatory requirements do not only define general risk management framework but also have significant impact on risk appetite definition, strategic decision making with special emphasis on financial transactions, stress tests and scenario analysis and risk reporting that encompasses all types and sources of risk to which bank is exposed. Krishna [10] states that requirements imposed on banks by regulators have brought transformational change to banks in regards to organizational structure, reporting hierarchies and on underlying data and technology infrastructure of the institution required to monitor risk and report it to the regulator. The reporting requirements are imposed not only on global-systematically important banks but also on banks of all sizes. One of the prerequisites for bank to be complied with regulatory requirements is banks' ability to apply a bank-wide risk governance system in order to achieve one single and truthful source of required information. When having all of the mentioned requirements in mind it is evident that banks need to use all the available data in order to predict risk, manage it and report it.

3.2. Impediments to big data usage in banks

One very important aspect of big data usage in organizations lies in managements' understanding that it is not the big data itself that has value but the ability of people in charge to extract value from the data for the organization. Bakshi [11] argues that "the challenge is not only to collect and manage vast volume and different type of data, but also to extract meaningful value from it". In order to use big data as a tool for business decisions support organizations are expected to incorporate big data analytics process. Gandomi and Haider [12] divide the process for extracting insight from big data into following sequences:

- Data Management (acquisition and recording; extraction, cleaning and annotation; integration, aggregation and representation) and
- Analytics (modelling and analysis; interpretation).

Power [13] defines three major types of analyses that can be prepared with new data sources and data manipulation technologies:

- *Retrospective data analyses* – using historical data and quantitative tools to understand patterns and results to make inferences about the future. This is the area of business intelligence.
- *Predictive data analyses* – using simulation models to generate scenarios based on historical data to understand the future. Predictive means ‘looking forward’ and making known in advance.
- *Prescriptive data analyses* – using planned, quantitative analyses of real-time data that may trigger events. Prescriptive analyses recommend actions.

Tian and al. [14] specify several challenges regarding big data analytics in financial organizations, namely:

- dealing with the problem of massive data storage and organization due to fact that many organizations need to keep historic data for many years for trend prediction and other complex analytics what poses a great challenge for reliability of the storage system,
- dealing with various data types since most of big data today is generated from different sources and thus unstructured,
- designing highly efficient computing system to process the distributed historical and incoming data,
- optimizing the usage of memory to keep the temporal data in memory is of special importance for risk management activities in financial organization due to requirements for fetching historical data in order to analyze new coming data.

In the context of stated that organizations need to use the analytics in order to make sense of diverse data, Capgemini Consulting [15] made a research which led to definition of three levels of building analytics maturity in organization whose characteristics are:

- Beginner:
 - There is no defined data structure/Informal, conflicting and dispersed data,
 - Poor data governance/Basic data reporting using mainly spreadsheet based tools,
 - Pockets of reporting and analysis capability/Dispersed talent,
 - Preliminary analytics strategy.
- Proficient:
 - Data available for existing and potential customers/Most data is still unstructured and internal,
 - Use of some statistical and forecasting tools/Coherent procedures for data management,
 - Well-defined recruitment process to attract analytics talent/Budget for analytics training,

- Analytics is used to understand issues and develop data-based options across the business.
- Expert:
 - Internal, external and social media data is merged to build and integrated and structured dataset,
 - Established robust master data management framework for structured and unstructured data sets,
 - Existence of analytics centre of excellence for promotion of best practice/strategic partnership for supplementary analytics skills,
 - Full executive sponsorship of analytics.

Evidently organisations on lower level of analytics maturity will have more difficulty in extracting value from big data for their internal processes.

According to a survey conducted by Capgemini and the Economist Intelligence Unit [16] biggest impediments for big data usage in effective decision-making process in organizations are:

- Too many “silos” – data is not pooled for the benefit of the entire organization;
- Time taken to analyse large data sets;
- Shortage of skilled people for data analysis;
- Big data is not viewed sufficiently strategically by senior management;
- Unstructured content in big data is too difficult to interpret;
- The high cost of storing and analysing large data sets;
- Big data sets are too complex to collect and store.

Duan & Xiong [17] state that big data contain more unstructured than structured data (namely text data, graph data and time-series data) and that biggest challenges for unstructured data are to:

- transfer unstructured data into a structured format,
- develop a new method to handle unstructured data.

According to [18] regular production of big data-based information takes time, especially because of resource issues. The resource issues in this context are related to material and human resources, i.e. big data analytics requires material investment, but above all it requires a multi-skilled professional who knows management needs and how to quickly get the desired reliable information from available large datasets.

4. Review of studies on big data usage in banks' risk management

In this section we will give a short overview of scientific and professional research on big data application in risk management. Up until now we have presented definitions of big data and their main characteristics. One of main questions banks' management is faced with is the source of big data. Based on the Oracle research findings [1] the authors made an illustration of possible sources of big data for banks shown on Figure 1. As can be seen the sources may be divided into internal and external, where internal resources encompass all kinds of data available from banks experts and applications and are usually “easy to grasp” for a bank from historical records, while external data

include some of the data in standardized formats (such as financial or trading data) but also some of more diverse and non-standardized data (such as social media data or alerts).

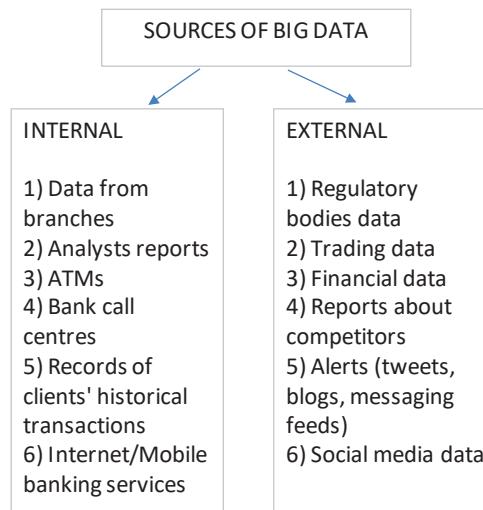


Figure 1. Sources of big data in banks, illustration by authors based on [1].

Part of the available literature explores different usages of big data in banks' risk management. Cerchiello et al. [19] used big data technology to propose a framework for bank risk contagion. The framework is based on graphical models that can estimate systemic risks with models based on two different sources: financial markets and financial tweets and suggest a way to combine them using Bayesian approach. This is the first systemic risk model based on big data and it sheds further light on the interrelationships between financial institutions. Ngai et al. [20] discuss application of data mining techniques in financial fraud detection, Ravisankar et al. [21] use big data technologies in detection of financial statement fraud and Hu et al. [22] use big data for the development of a network approach to risk management for modelling and analyzing systemic risk in banking systems. As stated in previous section regulatory requirements are one of the main drivers of big data adoption in banking industry. One of regulatory requirements all banks need to comply with is reporting. The subject of reporting is standardised and unique in desired outputs but every bank has the freedom to use different data sources and organize them in order to be compliant with regulatory reporting standards. Krishna [10] defines following enhanced data capabilities necessary for banks to be complied with regulatory reporting requirements:

- *Data sourcing* in timely, accurate and complete manner from catalogued sources,
- *Data processing and retention* should be efficient and able to support historical analysis,

- *Data analytics and reporting* should allow standardised and repeatable analytics with flexible drilldown and visual reporting capabilities,
- *Data management* with regard to access, retention, distribution and quality of data, data quality monitoring, error handling and reconciliation capabilities,
- *Data governance and control* relate to clear ownership, accountability and organisational standards combined with a robust data control framework.

The respondents of the Irving Fisher Committee *Report on Central banks' use of and interest in big data* [18] identified credit and market risk identification as an additional expected use of big data analysis. According to [23], over a half of risk managers stated that they lack sufficient data to support robust risk management. The Ernst&Young Global Forensic Data Analytics Survey [24] gives an overview of current big data usage in compliance monitoring:

- Payment stream and accounts payable analysis (altered invoices, duplicate or fake invoices, suspicious payment descriptions),
- Vendor master/employee master analysis and comparison (vendor risk ranking, conflicts of interest),
- Employee expenses/travel and entertainment (over limits, split or duplicate expenses),
- Payroll (ghost employees, falsified wages, commission schemes),
- Financial misstatement (fictitious revenues, concealed liabilities, overstated assets),
- Bribery and corruption (conflict of interest, bid rigging),
- Capital projects (contract non-compliance, project abuses and overcharges).

The results of [24] also suggest that 72% of respondents believe that big data can play a key role in fraud detection and prevention, 7% are aware of any specific big data technologies and only 2% are actually using them. When having in mind the growing complexity and number of financial transactions, volatility of markets, multiplication of devices, multiplication of channels and ever increasing demands from regulators, it is evident that what banks need is to make a connection between data acquisition, analysis and action [23]. McKinsey research [25] states that in near future some banks will lag behind because they are unable or unwilling to make investments into systems and infrastructure that will help them use big data for complex data investigation and analysis.

5. Formulation of proposal for strategic use of big data in risk management

When analyzing the state of current research that is available on the topic of big data usage in risk management in banks, we found that most of the existing research can be categorized in one of the following categories:

- drivers and challenges of big data usage in risk management in banks,
- practical application of big data in certain risk management processes (mostly modelling and fraud detection).

As mentioned in previous sections risk management plays a crucial role in contemporary banks' business activity. On one hand the regulator imposes strict rules banks must oblige in order to avoid penalty, but on the other in accordance with holistic view on risk management we believe that banks may achieve multiple benefits from adequate risk management. Some of these benefits include reduction of losses (costs) and enhancement of bank efficiency, enhanced management decision-making based on banks actual risk profile, building internal resilience and avoidance of systemic risk shocks. The foundation of adequate risk management, i.e. of any adequate decision management makes based on banks' risk exposure, is existence of reliable, truthful and comprehensive data.

We therefore propose a wider strategic look at the potential of big data in risk management by formulating a proposal on how big data can be used in different phases of risk management. In order to define this proposal, we conducted the following three steps:

- Selection of appropriate all-encompassing risk management system definition,
- Using this definition as basis for extraction of key risk management activities (phases),
- Allocating the potential usage of big data to key risk management activities defined in previous step.

Since the banking domain is subject to strict regulation the authors have decided to use the all-encompassing definition of risk management system from Basel III legislative as starting point for framework definition. According to [26]: "*Risk management system is made of organizational structure, rules, processes, actions and resources for identification, measurement or assessment, management and reporting on risk exposures, i.e. risk management in whole and it encompasses adequate corporate governance and risk culture.*"

As can be seen the used definition names several phases that are unique to every risk management process and these are: (I) identification, (II) assessment, (III) management and control and (IV) reporting. For the purpose of this paper we consider these phases key risk management activities. Final step in has been to allocate the known potential usage of big data to key risk management activities. In order to identify the potential usage of big data in risk management we have conducted a literature review and based on [23], [24] and [25] we have identified diverse usage of big data in management of various risk types (credit risk, liquidity risk, operational risk, market risk, interest rate risk in banking book). This list of usage is quite extensive and we have therefore extracted from it the usage that is common for all types of risks. This synthesized usage is listed in continuation:

- Early identification of new risk sources for bank,
- Integration of "know your client" principles with the process of early risk identification,
- Analysis of data underlying in calculation of various risk indicators,
- Real-time simulation of risk indicators,
- Predictive analytics for all risk types,

- Protection of bank brand via reputational risk management,
- Prevention of operational loss events,
- Compliance management,
- Real-time control of internal and external risk exposure limits,
- Real-time creation of reports,
- On-demand risk exposures calculation,
- Creation of reporting transparency,
- Real-time stress testing.

Figure 2 presents a scheme on how big data can be used in banks risk management on strategic level and is obtained by allocating the potential usage of big data to key risk management activities. As can be seen from the Figure, big data can indeed be used in every phase of risk management for any type of risk and may lead to enhanced risk management. Main argumentation for this statement lies in the fact that big data usage in risk management enables early stage reactions for banks management (presented as early identification of risks and real-time analysis and calculations). On the other hand, using big data as base for reporting and consequential increased reporting transparency increases banks good image in public what is of special importance from the agent-principal theory and postulates of reputational management.

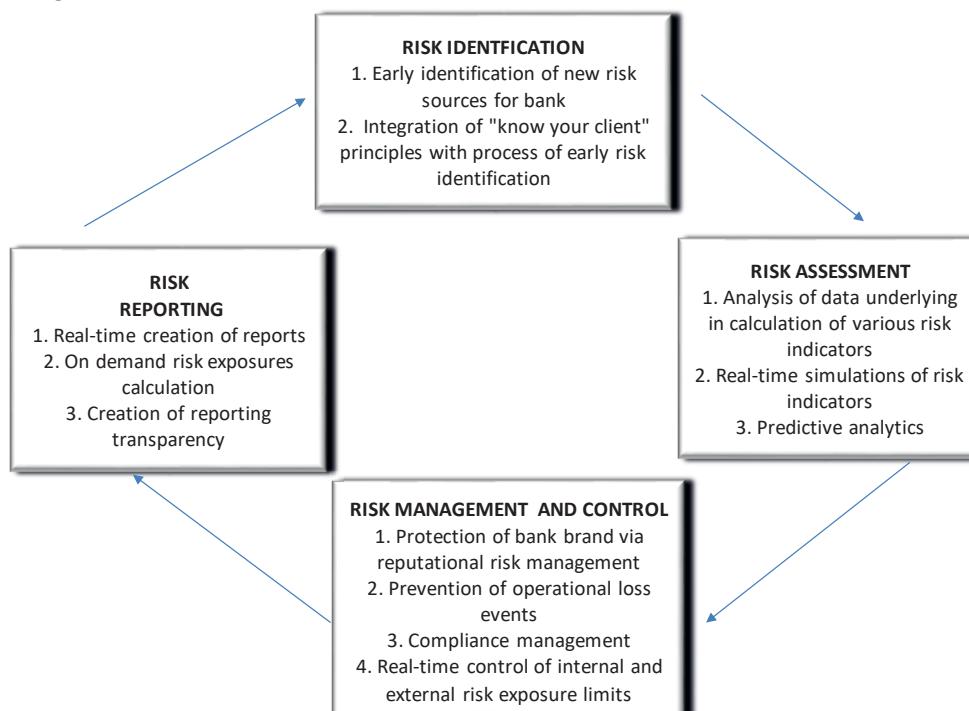


Figure 2. Big data usage in different risk management phases, illustration by authors

6. Conclusion

In this paper the authors have tried to present some aspects of big data usage in banks' risk management. Firstly, a review of available literature on the topic has been made. Based on it the authors have found that present body of literature is focused mainly on exploration of drivers and impediments to big data usage in banks' risk management or on practical application of big data in the risk management process. Also the importance of risk management in banks has been presented by giving an overview of existing regulatory requirements, but also by accentuating the holistic view on risk management in which risk is not seen just as negative aspect of business activity that needs to be managed just in order to avoid penalty, but also as an opportunity for a bank to achieve its strategic goals based on its actual risk profile. Therefore, the authors have presented a proposal on how to use big data in risk management for all the risk types banks are exposed to and in all phases of the process.

As mentioned the field of big data in banks' risk management is rather unexplored field in relation to other banking activities and therefore there is plenty of space for further research. Some aspects that deserve additional attention are: "on the field" research of big data practices employed in Croatian banks, exploration of their maturity levels according to big data usage, research on costs of big data technologies implementation and evaluation of cost/benefit ratio for big data technologies usage in risk management.

References

- [1] Oracle Enterprise: Big Data in Financial Services and Banking (2015)
- [2] Šverko, I. (2007). *Upravljanje nekreditnim rizicima u hrvatskim finansijskim institucijama*. Zagreb, Hrvatski institut za bankarstvo i osiguranje
- [3] Miloš Sprčić, D. (2013). *Upravljanje rizicima: Temeljni koncepti, strategije i instrumenti*. Zagreb, Sinergija nakladništvo d.o.o.
- [4] Gartner IT Glossary, What is Big Data? (2014), Retrieved from <http://www.gartner.com/it-glossary/big-data/> (17.05.2016.)
- [5] Deutsche Bank: Big Data: How it can become a differentiator (2015)
- [6] IBM: Operational risk management in the world of big data (2014)
- [7] Özköse, H., Sertac Ari, E., Gencer, C. (2015). Yesterday, Today and Tomorrow of Big Data. *Procedia – Social and Behavioural Sciences*, 1042-1050.
- [8] Mohanty, S., Jagadeesh, M. & Srivatsa, H. (2013). *Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics*. Apress.

- [9] McKinsey&Company: Digital and Big Data Opportunities in Credit Risk (2015)
- [10] Krishna, D. (2016). *Big Data in risk management*. Journal of Risk Management in Financial Institutions, 9 (1), 46-52.
- [11] Bakshi, K. (2012). Considerations for Big Data: Architecture and Approach. *Aerospace Conference IEEE, Big Sky Montana*
- [12] Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods and analytics. *International Journal of Information Management*, 35, 137-144.
- [13] Power, D. (2013). *Decision Support, Analytics, and Business Intelligence*. New York, NY: Business Expert Press.
- [14] Tian, X., Han, R., Wang, L. Lu, G. & Zhan, J. (2015). Latency critical big data computing in finance. *The Journal of Finance and Data Science*, 1, 33-41.
- [15] Capgemini Consulting: Big Data Alchemy: How can Banks Maximize the Value of their Customer Data? (2014)
- [16] Capgemini Consulting and the Economist Intelligence Unit: *The Deciding Factor: Big Data and Decision-making* (2012)
- [17] Duan, L.& Xiong, Y. (2015). Big data analytics and business analytics. *Journal of Management Analytics*, 2 (1), 1-21.
- [18] Irving Fisher Committee on Central Bank Statistics: *Central banks' use of and interest in "big data"* (2015)
- [19] Cerchiello, P., Giudici, P. & Nicola, G. (2016). Big data models of bank risk contagion. *DEM Working Paper Series*, 117.
- [20] Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y. & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50 (3), 559-569.
- [21] Ravisankar, P., Ravi, V., Raghava Rao, G. & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50 (2), 491-500.
- [22] Hu, D., Zhao, J.L., Hua, Z. & Wong, M.C. (2012). Network-based modelling and analysis of systemic risk in banking systems. *MIS Quarterly*, 36 (4), 1269-1291.
- [23] The Economist Intelligence Unit Limited: Retail banks and big data: Big data as the key to better risk management (2014)
- [24] Ernst&Young: Big risks require big data thinking, Global Forensic Data Analytics Survey (2014)

- [25] McKinsey&Copany: The future of bank risk management (2016)
- [26] Regulation EU No 575/2013 of The European Parliament and of The Council on prudential requirements for credit institutions and investment firms and amending Regulation EU No 648/2012 (2013)