BIG DATA: A Survey Paper on Credit Risk Management

SAHANA S KHAMITKAR
M.Tech II SEM (CSE)
Department of Studies in Computer Science and Engineering
University B.D.T College of Engineering, Davanagere - 577004, Karnataka, India
(A Constituent College of Visvesvaraya Technological University, Belagavi)

MOHAMMED RAFI
Department of Studies in Computer Science and Engineering
University B.D.T College of Engineering, Davanagere - 577004, Karnataka, India
(A Constituent College of Visvesvaraya Technological University, Belagavi)

Abstract

Big data is a buzzword that indicates data that do not fit traditional database structure. Their potential is enormous for many fields, and risk management is within the ones that could benefit the most from new sources of unstructured data. This paper introduces the big data framework, terminology, and technology, in order to understand the upsides and challenges that they pose to financial markets. A review of standard methods and tools in risk management is then provided, in order to be able to understand the revolution brought into the environment by big data. Simulation and forecasting are the two areas that are affected the most, and therefore the ones of interest for this study. Nowadays banks operate in changing environment influenced by regulatory requirements, emerging risk types and competition on the market. At the same time banks have available large datasets arising from internal and external sources. The potential for this data usage in risk management has only recently been discovered and has not been the subject of extensive scientific research. There are two goals of this paper. Firstly, authors give an overview of available scientific literature and practical research related to big data usage in risk management in banks. Secondly, based on the literature review authors are presenting framework with specified detailed use of big data in specific risk management area.

Keywords: *Big data, Banks, Risk management.*

INTRODUCTION

Since the 1990s, the increase in competition worldwide and the availability of structured and unstructured information have deeply changed companies production and organizational processes, making it necessary to improve upon traditional management and data analysis systems **[10]**. To extract value from information, it is necessary to consider the systematization of knowledge through the creation of archives and the management and analysis of data produced in large quantities (volume), with rapidity (speed) and in different formats (variability). This was the starting point for the evolution of business intelligence through the implementation of software dedicated to information processing, services to integrate new technologies with existing systems, and infrastructural resources with an increase in computing and storage capacities for data from which to derive new levels of knowledge (Big Data analytics [BDA]).

This new paradigm in the digital era requires the support of multiple skills and the knowledge of programming techniques and allows a better interpretation, inspection, cleaning and modelling of large amounts of data extracted from various sources (including the web) that are useful in the decision-making process The data coming from the web plays a crucial role in the context of Big Data, considering its high information potential, especially in forecast analysis. Operational data sources include accounting applications and personnel and customer management, which are appropriately integrated with other segments, such as production management; purchasing and delivery applications for companies; industrial and branch management; financial instruments; and risk assessment for banks.

In recent years, the process of sharing information both internal and external, and particularly concerning risk,

governance and performance, is increasingly supported by the evolutionary paths of new technologies, creating the premise for a gradual revolution of business and organizational processes and placing itself as the foundation of the 'fourth industrial revolution'. The collection of large amounts of data in a heterogeneous, redundant and unstructured form (Big Data) and its interpretation, analysis and evaluation mainly follow risk-based logics that are primarily focused on credit, operational and compliance risks (e.g., the GDPR directive and anti-money-laundering legislation). The importance of Big Data emerges most notably in the financial, banking and insurance sectors due to their supervision authorities which impose stringent capital regulations.

The amount of data acquired through digital technologies and multi-channelling with the adoption of BDA could support the maximization of global business value thanks to the alignment of strategic priorities for risk management activities, the timely reporting of sources of uncertainty on which to focus attention, and the implementation of specific actions to improve performance. The knowledge and measurement of risk and the subsequent identification of anticipatory and proactive actions constitute priority activities that affect the achievement of strategic objectives through the assignment of specific responsibilities to all company levels and the creation of an efficient reporting and communication system. The support of communication (top-down and bottom-up) within an organization is fundamental, as is the proliferation of relevant information for the census and filing of all external risks to business processes and information for decision-making.

Why big data are relevant to risk management:

The exponential increase in the amount of data created in the last few years impacted every sector. Big data, i.e., datasets so large to not be eventually fit into traditional databases, revolutionized the way we deal with every decision making process, as well as approach to several business and research questions. Even though the list of innovative applications can fill many pages, one of the greatest changes brought by big data concerns financial markets. For example, many works have been implemented using sentiment analysis, i.e., whether a piece of text reflects a positive/negative sentiment**[1][3][2]** and several algorithmic trading companies have been started with the help of massive datasets. Hence, even though many innovative steps have been undertaken in financial markets, not all the categories have been affected in the same way: in particular, new information and stack of technologies did not bring as many benefits to the risk management as they did to credit fraud detection for example.

Risk is usually addressed from an operational perspective, from a customer relationship angle, or with a specific focus on preventing fraud and credit scoring. Greater uncertainties, linked to political, regulatory, macroeconomic and technological factors, have ensured that risk management activities have taken on both increasingly broader connotations and a continuous and integrated approach aimed at mitigating risky events with widespread responsibility inside a company's organization. The implementation of adequate risk management policies and programmes requires the involvement of the highest levels of the corporate hierarchy to define the essential principles for protecting internal and external subjects (social responsibility) and safeguarding business continuity and generating value over time (behavioral ethics). The rethinking of organizational models to focus them to a greater extent on a systemic view of corporate risk was inevitable. An important contribution to the formalization of the new risk management approach has been provided by several standards, of which the most complete are the International Organization for Standardization (ISO)/IEC 31000 Risk Management principles and guidelines and the Committee of Sponsoring Organizations (CoSO of the Tread way Commission) Enterprise Risk Management integrated framework, which include principles and guidelines for the integration of the risk management process into the overall governance of an organization, from the strategic planning process to the reporting policies. The present work investigates the level of implementation of advanced technological infrastructures adopted by small banking institutions and their ability to identify and create an effective risk management process in light of the limits established by internal regulations (risk appetite) and supervision. investigated in the literature, despite risk management representing an area of great importance, given the central role that banking intermediaries have in the economic fabric and the consequences of their bankruptcy.

However, applications strictly related to financial markets are still not so common: even if in theory more information should entail a higher degree of accuracy, in practice it also exponentially augments the system complexity, and makes extremely hard to identify and analyze timely unstructured data that are valuable. The Deposit Trust Clearing Corporation had recently identified many potential risks in financial markets, and a vaster amount of data can help institutions and banks in addressing them: high frequency trading risk, liquidity and credit risks, collateral risk, counterparty risk, only to name a few. Markets are always more interconnected, which also increases the risk of a network systemic failure: more and more data can help central institutions and regulators to predict in real-time symptoms of a future
crisis, and acting on time to prevent it or weaken it.

A particular field of interest regards the market and counterpart risks. The huge complexity introduced in the market made the common pricing techniques obsolete and slowly reactive, and required a more comprehensive pricing approach than a net discounted value of derivative's legs. This is the reason why banks and financial institutions need (but struggle) to simulate a

single portfolio a hundred thousand times, or because an accurate fast forecast is considered a breakthrough achievement.

The purpose of this paper is to give an overview of common risk management models, and of how simulation and forecasting models are modified and improved by a larger amount of data.

Big data technologies:

Technology plays an important role into this field, and the transformations we have taken part of are outstanding: the data traffic has been redirected to the cloud, through a shared pool of connected storage devices; parallel computing method helped in computing not use structured query language and allows to consider data that would not usually fit a standard table.

Those technologies allow for improvements in timing and predictive intelligence into the risk management field. The degree of innovation these tools and techniques are bringing is remarkable, especially when it comes to real-time simulation and large– volume forecasting.

Traditional risk management methods:

We briefly survey the most common techniques and methods contained in a risk manager toolbox. The initial approach to begin with is called Value-at-Risk (VaR), which is used to assess the market risk. In a nutshell, the VaR is a statistical technique used to measure the level of risk of a portfolio given a certain confidence interval and within a fixed time frame. The VaR is a threshold value such that the potential loss over a specified time period is equal to a given probability. In other words, given the confidence level α, the VaR is that number k that makes the probability of a loss L greater than k be exactly equal to $1-\alpha$:

$$VaR_{\alpha} = inf\{k \in \mathbb{R}: \Pr(L > k) \leq 1-\alpha\} \qquad (1)$$

Many different variations have been proposed over the years, and a particular attention has to be devoted to two coherent risk measures alternatives, i.e., the conditional VaR (CVaR) and the entropic VaR (EVaR). The CVaR, also called expected shortfall, indicates for a certain probability level the expected return of the portfolio in the worst scenarios:

$$CVaR_{1-\alpha} = \frac{1}{\alpha}\int_0^{\alpha} VaR_{1-\gamma}(X)\,d\gamma \qquad (2)$$

The EVaR instead represents the upper bound for both VaR and CVaR, and its dual representation is related to the concept of relative entropy:

$$EVaR_{1-\alpha} = inf_{z>0}\left\{\frac{1}{z}\, ln\frac{M_L(z)}{\alpha}\right\} \qquad (3)$$

where $M_L(z)$ is the moment-generating function of the loss.

larger amount of data, reaching a greater accuracy for the same costs; and finally, unstructured data have been taken into account, thanks to the implementation of new developments in the data architecture space. Few technologies seem to be commonly used across financial markets player: Hadoop in undoubtedly the most used platform for managing unstructured data in a parallel computing setting. Along with MapReduce, it allows for an efficient data cleansing and massaging, as well as for complicated or CPU-demanding computations: the task is indeed split between many servers (independent machines with their own memories and operating system), computed locally, and the results are reaggregated afterwards. A second important tool in a non-relational database named NoSQL, which does

The VaR, regardless of the type, is usually computed through either historical method, the Delta-Normal one, or Monte Carlo simulation. The first method just lists historical returns in ascending order, while the Delta-Normal technique looks back in the time series, computes mean, variance, and correlation, and finally obtains the portfolio risk through a combination of linear exposure to factors and the covariance matrix. The last method, i.e., the Monte Carlo simulation, is probably the most used nowadays, as well as the most interesting from a big data perspective. It actually requires to develop a model for the stock price/returns trajectories, then runs a multitude of simulated trials and averages the results obtained. The Monte Carlo is then a repeated sampling algorithm that could be used for solving any problem that may be stated through a probabilistic lens, and which is often exploited for pricing extremely complex derivatives.

More recent advances in risk management tools are related to credit counterparty risk. The framework known as the X-Value Adjustment (XVA) includes credit valuation adjustment (CVA), debt valuation adjustment (DVA), and funding valuation adjustment (FVA), and respectively deals with the risk of the counterparty, the risk of the entity itself, and the market value of the funding cost of the instrument.

$$CVA = E_0\left[\sum_n \int_0^{\infty} e^{-\int_0^t r_s ds}\,(P_t^n) + d\pi_t^n\right] \qquad (4)$$

where P is the price process of the portfolio, n is the index for netting sets, r is the short rate and π is the process of the cumulative probability of default. The intuitive interpretation for the CVA is that it represents the market value of counterparty credit risk, and it is obtained as the difference between the risk-free portfolio and the portfolio that embeds a potential counterparty's default.

The problem with XVA measures is that they require a huge amount of computation power to be calculated effectively. Even though for a standard portfolio CVA calculation a reasonable number of simulations are required, banks might need to run many more deals if they want to take into account all the path-dependent derivatives in their portfolios - the number would be indeed close to several hundred thousand simulated paths.

Big data methods:
The econometrics of large datasets:

Big data are everywhere. Social media provide for example an endless source of information for financial market, because the market moves following the actions of the crowd. and the different data affect different risks with a distinctive intensity. The following table summarizes their findings rating from 1 (the feature with the strongest impact) to 4 (the weakest benefit) the impact of each characteristic on each risk (each cell is evaluated independently from the others):

Risk Area	Volume	Velocity	Variety	Veracity
Credit Risk	1	4	3	4
Market Risk	3	3	4	4
Operational Risk	3	4	4	3
Compliance	2	3	2	2
Asset-Liability risk management	2	4	3	4

Table 1: Impact of big data features on risk management

The availability of those endless data cannot solve every single problem, and in fact big data poses as many technical challenges as well as opportunities for organizations and regulators , such as lack of skills, issues related to hypothesis/testing/model, or hardware/software challenges. The main challenge the increase of noise into the signal ratio, to the detriment of the actual predictive power of the additional data. The forecasting techniques then have to be able to filter down that noise and leave the model with only the variables and data that matter and at the same time to provide accurate out-of sample forecasts without abusing of a large number of predictors. In addition, conventional statistics techniques face two additional issues when big data are added to the equation: a higher degree of data manipulation is required, because every data problem is exponentially amplified, and large data allow for different relationships than linear ones.

Hence, the goal of the next two subsections is to provide a summary of models that prevent over-fitting and that are able to manage large datasets efficiently. In general, simpler models work better for out-of-sample forecasts, and excessive complexity should be avoided.

Big data simulation:

Simulations considering huge data amounts allow for an efficient realization of risk concentrations and quicker reactions to new market developments. In particular, Monte Carlo simulation is a powerful and flexible tool, and the challenge with that is finding the optimal number of paths to match speed and accuracy. A higher accuracy is achieved by the larger amount of simulation the model can project, but it has been always bounded by a lower processing speed as well as machine memory. Even though a set of techniques have been used to handle this burden, the only solution lies in splitting the data between many different workers.

Parallel computing is gaining popularity, and many algorithms for making it less expensive have been developed in the last few years**[9]**. Hence, two main methods may be used in order to relief a single terminal from a great data burden: either it can be divided into different cores on the same chip, or it can be divided through different machines. In the first case, the splitting can be made on multi core CPU, or on parallel GPU**[9]**. In any of the two cases, few problems arise: difficulty in writing the splitting configuration, absence of positive effect on memory, and difficulty in abstraction make those methods cumbersome to be used. The second alternative instead is much more scalable: dividing data into different machines increases the processing power and efficiency, although it comes with a higher cost. A solution to this problem has been proposed in **[9]**, called consensus Monte Carlo: this new model runs a separate Monte Carlo algorithm in each terminal, and then averages individual draws across machines. The final outcome resembles a single Monte Carlo set of simulations run on a single machine for a long time.

Big data forecasting:

Many completely new methods from one hand, and techniques borrowed from other disciplines from the other, are nowadays available to researchers and practitioners. classification of forecasting methods into four groups: single equation models that use the whole datasets for estimation purposes; models that use only a subset of the whole database, even though a more complete set is provided models that use partial datasets to estimate multiple forecasts averaged later on in a conclusive result and finally, multivariate models that use the whole datasets with the aim of estimating a set of variables.

The first group is quite wide, and includes common techniques used differently, such as ordinary least square (OLS) regression or Bayesian regression, as well as new advancements in the field, as in the case of factor models.

In the second group, we expect the most of the data to be noisy and not really meaningful, thus we use features or variables selection models to identify ex-ante the more significant predictors. This class of methods provides then a stratagem to avoid, if wanted, the problem to deal with large datasets.

On the other hand, the third group is the averaging model one, which embeds mainly the Bayesian model averaging (BMA) and frequentist model averaging. While the frequentist approach entails the creation of model confidence sets from which model likelihood can be derived, the BMA methodology provides a Bayesian framework for the combination models. Different

combinations of relationship between predictors and dependent variables are estimated, and weighted order to obtain a better forecasting model, with weights corresponding to the posterior probabilities.

Finally, the last group uses the whole datasets to estimate a set of variables **[3]**. Reduced rank regression, Bayesian VAR, and Multivariate Boosting. Reduced rank regression works similarly to a set of classic Vector Autoregressive models, but as soon as underlying dataset is large, those models become quite noisy and rich of insignificant coefficients. Therefore, a rank reduction can be imposed, constraining the matrix of the coefficients in the VAR model to be much smaller than the number of predictors. Differently from this model, which compressed the greatest informative value into few predictors, the Bayesian counterpart of the VAR focuses instead on constraining the data imposing the restrictions as priors, although it maintains a dependency between data and coefficient determination. On the other hand, the multivariate boosting is quite similar to its simple version explained beforehand. The main difference lies in measuring at each step a multivariate model (instead of a single equation as in simple boosting), starting from a zero coefficient matrix and recursively setting the single coefficients that explained better the dependent variables to be non-zero. Afterwards, a second prior is selected, i.e. a prior on the regression coefficient deriving from the previous choice of certain variables.

Framework for big data usage in risk management in banking:

According to Basel III regulatory requirements risk management control function, together with the compliance function and internal audit function, presents one of the three bank's control functions. Main tasks of control functions are insurance that the bank is conducting its business activity in accordance with defined strategy, rules and policies on risk management. The findings of risk management control function are subject of internal and external reporting. Internal reports are related to management board and supervisory board reports, while external reports are subject of central banking authority monitoring. Therefore, the findings of control functions incorporated in reporting form basis for bank's management strategic business decisions, but also a tool for external supervisory monitoring. Having on mind the role and importance of risk management in individual bank on impact on stability of altogether in financial sector in whole, it is evident that banks need to adequately manage its risks and exploit all the possibilities, tools and technology that they have on their disposal in order to adequately manage its risks.

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:

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

When having in mind the importance that risk management has for banks we propose a wider strategic look at the potential of big data in this process by defining a framework on how big data can be used in the risk management process. There are several steps that we conducted while defining this framework for all-encompassing big data usage in banks risk management:

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

following key risk management activities have been identified:

1) identification,
2) assessment,
3) management and control and
4) reporting.

Final step in framework definition has been to allocate known potential usage of big data to key risk management activities and it resulted in the formulation of the proposed framework that we are proposing in continuation. The framework is shown by Figure 1. As can be seen from the Figure, big data can be used in banking in all key risk management activities in order to enhance the Process of making various types of risks banks are exposed to.

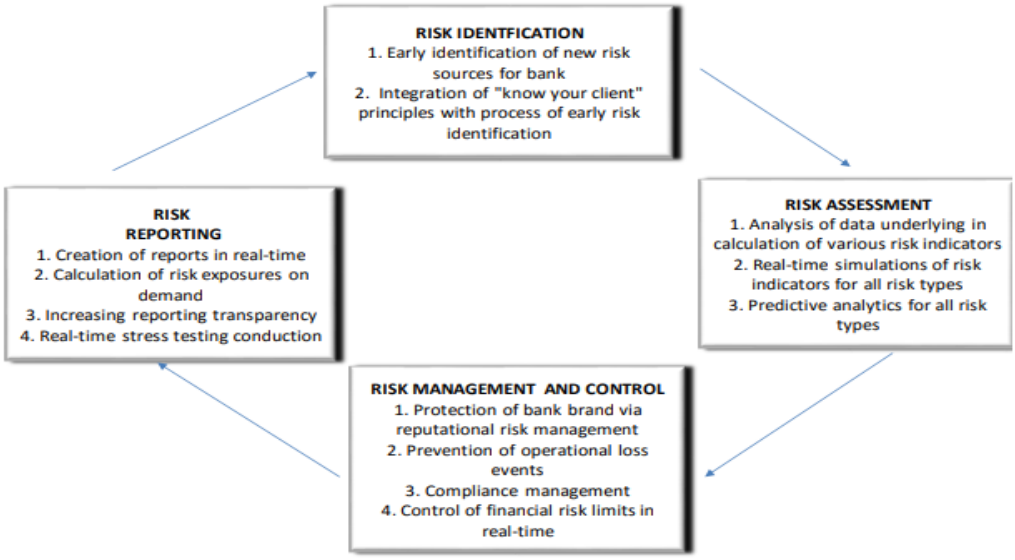

Figure 1: *Framework for big data usage in risk management*

LITERATURE REVIEW

The Role of Big Data in the Digital Economy:

Big Data already existed at the end of the 1990s and has spread enormously in the 21st century, becoming, in the current context, a key element for modern business. Companies all over the world are exploring these large volumes of highly detailed data to discover previously unknown information that is useful in improving the decision-making process[**5**]**.** Big Data refers to sets of data so large that they cannot be used with traditional database management systems because their dimensions overwhelm the capability of the software tools and storage systems commonly used to acquire, store, manage and process data within a tolerable time frame[**5**].

The literature identifies three main features that characterize Big Data, also known as the 3Vs:

1) Volume,
2) velocity and
3) variety[**8**].

Volume refers to the quantity of data and, therefore, to the dimensions of the dataset. Velocity refers to the speed of the data flow, that is, the rate at which the information is generated and spread and therefore the rate at which it is processed and analyzed (velocity of data and processes). Variety is the characteristic that makes the data 'big' and is related to the typology of the information sources and the generated data, which can be structured, unstructured or semi-structured and can derive from different sources, both internal and external. Some researchers attribute to Big Data two more characteristics: Variability and veracity. Variability concerns the periodicity or irregularity and, sometimes, the incoherence of the. Veracity concerns the accuracy of the data, which can be good, not good or undefined, with the data potentially being incoherent, incomplete or ambiguous. Some authors identify a further characteristic, value, referring to the potential value of the data[**8**]**.**

In 2015, the United Nations Department of Economic and Social Affairs classified Big Data into three categories according to the different sources from which it derives: Data from social networks, including information from social media, messages and research conducted on the internet; data from traditional systems of business, such as that generated by commercial trade transactions, e-commerce, credit cards and medical records; and data from the so-called Internet of Things (IoT), referring to machine-generated data, such as that concerning weather and pollution, data from GPS satellites and data from computer-based registers[**5**].

The process of extrapolation of information is articulated in two phases: The first, known as data management, consists of the acquisition, storage, selection and representation of data; the second, called analytics, is composed of all the activities focused on the analysis and interpretation of data. Data is first extracted through information system tools from external sources, then transformed and loaded into advanced databases or data warehouses. The data can then be cleaned and classified before making it available for data mining and other forms of analysis. Finally, it is processed and submitted to the BDA tools necessary to make the Big Data useful in the decision-making process.

Many researchers note that technologies relating to Big Data are applicable in many areas of the banking sector, including retail (bank collections, credit cards, private banking), commercial (credit risk analysis, customer and sales management, middle market loans), capital markets (negotiation and sales, structured finance) and asset management (wealth management, management of capital investments, global asset reporting, analysis of investment deposits)[**7**]**.**

Risk Management in the Big Data Era:

The correct functioning of a business activity and the contextual creation of an enterprise's economic value cannot neglect monitoring of the main risk factors, as represented by financial and managerial indicators whose economic effect can compromise performance. The clear and evident interconnections and interdependences between business risks have led to an increasingly global management of enterprise risks following a systemic approach that is coherent with the growth path of a company and a contextual transversal analysis of heterogeneous processes, functions and activities. It follows that, in the last few years, overhauling of the traditional approach characterized by a mainly sectorial and fragmented view of risks ("silo" management) has resulted in the spread of a new philosophy in the management of business risk that involves the whole organizational structure and affects strategic and operational processes. This approach is known as enterprise risk management (ERM) and provides for integrated risk management through an analysis of business contingencies and an evaluation of uncertainty, with organizational solutions recognized and shared by the whole company, with the aim of business continuity.

At the operational level, important international contributions have come from ISO and CoSO, which have outlined a series of principles and operational techniques for a more systematic and disciplined approach to risk management through the interaction of adequate control systems with performance and business strategies. These strategies, even if they are in a state of uncertainty, can lead to the creation, maintenance and realization of value and, hence, to the satisfaction of the stakeholders' expectations for long-term sustainability. In recent years, the integrated risk management approach has been implemented by a growing number of companies in light of its competitive advantages, its increase in the economic value of business capital, the improvement of operational performance and the reduction of the risk of failure and in the awareness that 'in order to create and protect the value of an organization it is essential to manage the risks in a structured way and based on well-defined principles.

It is obvious that banks increasingly need to use all available data to predict risks, manage them and report them. The quantity and quality of data are essential elements for the formulation and implementation of strategies compatible with risk appetite and suitable for structuring effective and reliable processes and procedures for safeguarding the integrity of bank assets. This has not only required a revision and adaptation of organizational models, but also emphasized the importance of technology in both the automation and integration of the various processes (stress testing, modelling, monitoring, reporting, capital planning, etc.) and in the management of big volumes of irregular data from which, by means of advanced analytics systems, it is possible to draw information useful to the management of organizations and processes. According to research conducted by Deutsche Bank (2015) and McKinsey and Company (2015), the increase in banking tools and transactions, the subsequent explosive growth of data, the increasingly innovative use of statistics and mathematical sciences of risk management, the development of new typologies of risks and, above all, the increasingly stringent regulation and attention have had an immense impact on the data underlying the information and technological infrastructure that play an increasingly central role in the value creation process. The intent is to ensure that banks can integrate traditional data coming from various channels traditional counters, internet and mobile banking, ATMs, ASDs/ASSDs (assisted self-service devices), credit circuits and e-commerce platforms with unstructured data, especially from social networks and the web, through the creation of a data lake. This is a platform that allows storage, organization, management and operational exploitation of large volumes of data to return helpful information and knowledge of the consumption habits and risk profiles of customers for fraud.

Many authors support the idea that rapid identification and quantification of new risks and a transparency in reporting activities are essential in risk management **[7]** to integrate traditional information sources with unstructured data acquired from various internal and external sources (Word, Excel, PowerPoint, images, e-mails and information from the internet) using advanced technological tools and new data-intensive techniques for the construction of a shared platform the so called BDA**[7]**. These advanced tools include data sourcing tools that find data in a timely, accurate and complete way; data processing and retention tools that process and store data in an efficient way and support historical analysis; data analytics and data reporting tools that conduct advanced analyses and detailed reporting; data management tools that manage access, storage, distribution and quality of data; and data governance and control tools that govern and control data with reference to property, responsibility and organizational standards concerning usability, accessibility, accuracy and consistency of data.

[7] develop a framework in which they suggest the use of Big Data in each of the four key risk management activities (identification, assessment, management and control, and reporting). The framework can be articulated as follows:

1) Risk identification: Identification of new sources for the early identification of risks and in-depth knowledge of customers.
2) Risk assessment**:** Analysis of underlying information through the calculation of various risk indicators, real-time simulation of risk indicators and predictive analysis for all typologies of risk.
3) Risk management and control**:** Reputational risk management, operational loss forecasting, compliance management and real-time control of financial risk.
4) Reporting: Real-time creation of reports, calculation of risk exposure on request, increased transparency and real-time stress tests.

The literature shows how Big Data can be a crucial element in risk management systems, especially in predictive analysis of credit institutions. Large banking groups have an organizational structure that can support innovations in the IT field, but the collection and management of data by smaller banks remains an open question. Starting from this premise, the present work intends to answer the following research questions:

1) What are the techniques of collection and processing of data currently employed by small banks in the area of risk management?
2) How can data management technologies affect the risk management process?
3) How will the use of Big Data affect the risk manager's skills?

METHODOLOGY

To answer the research questions, a case study methodology was used **[11]**, as it was considered particularly suitable for observing a complex phenomenon **[4]**, such as the one in question. The survey was conducted using a single case study a credit institution presenting the typical characteristics, in terms of organizational structure, of a small bank. Based on the objectives of the research and the current phase of the development of Big Data in risk management, the analysis of the case study was of an exploratory nature, to provide preliminary explanations for the research questions, which will need to be expanded through subsequent empirical investigations.

A semi-structured interview with open answers was conducted with the credit institution's head of risk management, to understand the effective use of Big Data in the risk management function. The interview, lasting about 60 min, was conducted at the bank's headquarters, which allowed observation of the organizational structure in terms of both human and technological resources. The interview was structured based on three survey profiles:

1) Risk management and information technology actually used in risk management;
2) Future perspectives about the use of Big Data; and
3) New risk manager skills.

CONCLUSIONS

In this paper a review of most common concepts and techniques in risk management has been provided, and then an extensive list of models useful in a big data setting has been made available. Although is not possible to identify an algorithm that works the best in every situation (a kind of fit-them-all procedure), a combination of different models can be used efficiently to increase the model accuracy, reliability, and practical utility. Big data are clearly not a panacea for all the business and markets issues, but they offer for sure a new view on things, enlarging the spectrum of relationship we are able to analyze, providing new horizons and directions for professional to act in a more efficient, ethic, and compliant way, and increasing our knowledge of the markets themselves.

In conclusion, the challenge for the banking sector is to import non-conventional technologies (IoT, business intelligence, Big Data and blockchain) that allow the processing of huge quantities of data quickly and accurately, which is useful not only in CRM, but also in cyber security, fraud detection and the optimization of decision-making processes. A technological innovation must be accompanied by a cultural change and, in this case, the formation of teams composed of statistical, mathematical and technological skills and data scientists who can combine data analysis skills with functional skills to create automatic processes of value. This ongoing disruptive change necessarily affects people and involves the whole organization, including IT, marketing, business and management control, strategic planning and customer service, to obtain significant advantages in terms of risk analysis, fraud prevention and advanced analysis of customer intelligence through the storage, organization, management and operational use of large quantities of data; this will be part of a future evolution. The knowledge potentialities of BDA are therefore accompanied by a series of criticalities, starting with the risks of data confidentiality and progressing to organizational problems connected to the need to cooperate with those who can use the algorithms and reorganize internal information systems.

REFERENCES

1) **[1]**Bollen, J., Mao, H., Zeng, X. (2011). "Twitter mood predicts the stock market". *Journal of Computational Science Volume 2 (1): 1-8. 5.*

2) **[2**]Corea, F., Cervellati, E. M. (2015). "The Power of Micro-Blogging: *How to Use Twitter for Predicting.*

3) **[3]**Carriero, A., Kapetanios, G., Marcellino, M. (2011). "Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models". *Journal of Applied Econometrics 26 (5): 735-761. 6.*

4) **[4]**Eisenhardt, K.M. (1989), Building theories from case study. *Academy of Management Review 14(4), 532-550*.

5) **[5]**Hasnat, B. (2018), Big data: An institutional perspective on opportunities and challenges. *Journal of Economic Issue, 52(2), 580-588.*

6) **[6]**Kroese, D. P., Brereton, T., Taimre, T,, Botev, Z. I. (2014). "Why the Monte Carlo method is so important today". *WIREs Computational Statatistics 6: 386-392.*

7) **[7]**Lackovic, D.I., Kovska, V., Lakovic, V.Z. (2016), Framework for Big Data Usage. Risk Management Process in Banking Institutions. *Central European Conference on Information and Intelligent System. p49-54*.

8) **[8]**Ozkose, H., Ari, E.S., Gencer, C. (2015), *Yesterday, today and tomorrow of big data. Procedia Social and Behavioural Sciences, 195, 1042-1050*.

9) **[9]**Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H., George, E., McCulloch, R. (2013). "Bayes and big data: The consensus Monte Carlo algorithm". EFaBBayes 250 conference 16.

10) **[10]**Vasarhelyi, M.A., Kogan, A., Tuttle, B.M. (2015), *Big data in accounting: An overview. Accounting Horizon, 29, 381-396*.

11) **[11]**Yin, R.K. (2014), Case Study Research: Design and Methods. *Thousand Oaks, CA: Sage Publications*.