THE BENEFITS OF BIG DATA AND ADVANCED ANALYTICS IN BANKING SYSTEMS IN CONTEMPORARY ENVIRONMENT

Vladimir Mirković[1]*,
[0000-0002-0550-211X]
Jelena Lukić Nikolić[2]
[0000-0003-0632-8974]

[1]Economists Association of Belgrade, Belgrade, Serbia

[2]Modern Business School, Belgrade, Serbia

Abstract:

Business processes in organisations have changed dramatically with the advent of digitalisation and are facing numerous challenges in the age of Big Data. This new era is characterised by the presence of traditional banks and Fintechs that can cooperate or compete with each other. Banks and Fintechs have access to large amounts of customer data that could significantly improve the decision-making process. Access to external and unstructured data is the basis for better decisions in assessing customers' creditworthiness and credit risk, but also for monitoring the quality of bank portfolios and developing early warning systems, detecting financial crime and predicting operating losses. Advanced Analytics looks at the use of machine learning and its algorithms when processing large volumes of data. Machine learning has become a driving force in the banking sector and is driving improvements in decision-making processes. Machine learning algorithms can be used to fulfil regulatory requirements in a more sophisticated way, but also in the "front office", "back office" or in trading and portfolio management. The paper emphasises various possible applications of machine learning in the banking sector, with a focus on the area of risk management.

Keywords:

Big Data, Advanced Analytics, Machine Learning, Risk Management, Digital Transformation.

INTRODUCTION

In the banking sector today, a mixture of traditional data that companies collect about their customers, suppliers, products and services, with Big Data and third-party data containing additional information, such as demographic and geographic data, is used. Patterns could emerge from the data collected, providing a fundamental analytical tool for predicting customers' future behaviour. The importance of data obtained through the exchange of information in social networks is also constantly increasing and contributes to the development of high-quality techniques for predicting customer behaviour.

With the strong penetration of digitalisation into all areas of society, business processes in companies have changed significantly. Banks, as financial intermediaries between companies with surplus and deficit funds, are creating their market position in the modern era by adapting to all the changes required by the new Big Data era.

Correspondence:
Vladimir Mirković

e-mail:
vladamirkovic@orion.rs

DOI: 10.15308/Sinteza-2024-122-127

In this sense, every business process in modern banks runs according to the laws of digital business, with the inevitable adaptation of business processes to the banks' primary lucrative goals. The introduction of new technical solutions to optimise and modernise business processes leads to a greater focus on certain types of operational risks, such as cyber risks, fraud, anti-money laundering and combating the financing of terrorism. The application of Big Data contributes significantly to the improvement of risk management in banks (e.g. through the application of the scorecard system), as management level initiatives can be interpreted much faster as a direct result of the increased availability of relevant information [1].

In addition to banks as traditional players in financial systems worldwide, the last decade has been characterized by the development and “rapid" rise of Fintechs and their role. This has given rise to the dilemma of whether banks and Fintechs will act as competitors or as co-operation partners in the reshaped financial world. Empirical examples of collaboration between Fintechs and traditional banks are reflected in the establishment of different types of digital banks, such as neo-banks, new banks, beta-banks and non-banks. At the same time, some banks such as Singapore-based DBS Bank Ltd. have taken a leading position in the digital banking market and benefit from the integration of modern digital solutions into their business [2].

Apart from the fact that banks are financial entities that have greater financial strength and an impressive market share in the financial markets, while Fintechs (usually new, start-up companies) are trying to occupy market niches abandoned or forgotten by banks, it should be noted that Fintechs derive their competitive advantage from their agility, flexibility and innovation. In the future, it is more realistic to expect that market-orientated banks will choose to collaborate with Fintechs as they are able to leverage banks' existing infrastructure, reputation and image with customers and their larger customer base to offer innovative products and services [3].

The article is organized as follows. After the introduction with a description of the most important relationships between banks and Fintechs in the modern Big Data era, the second title focuses on the main characteristics of Big Data, which are reflected in the “7V” model. The third title presents the key findings from the European Banking Authority's (hereinafter: EBA) report on Big Data and Advanced Analytics in the banking sector. The fourth section covers the most important developments in the field of risk management, in particular in the area of machine learning (hereinafter: ML) used in an internal ratings-based model (hereinafter: IRB). Finally, the main findings and recommendations for future research are summarized. The next section is dedicated to the topic of Big Data and its main characteristics, which are described with concepts from “3V” to “7V”, as well as its implementation in the banking sector.

2. MAIN CHARACTERISTICS OF BIG DATA: “7V” CONCEPT

The amount and type of data that companies and financial institutions use in their business has grown rapidly over the last decade. Before the advent of Big Data technology, banks were not able to process and analyse all the data available to them in the right way and in real time. The application of Big Data technology and working with large amounts of data in real time enabled the development of new, creative solutions and the establishment of new business models. In addition, numerous potentials for value creation were observed through the combination and integration of internal and external data from different sources. Big Data usually refers to the concept of “3V”, which comprises the following characteristics: Volume (amount of data), Variety (data from different sources) and Velocity (the speed of data movement) [4]. At the centre of banks' information technology systems is the core banking system, which is connected to large amounts of data, usually covering deposit, loan and credit processing. The data in the core banking system can come from various sources (internal or external) and should be organised efficiently so that it is useful to users. The aspect of speed in the financial industry is given by the availability of customer data in real time. The provision of data in real time is a prerequisite for modern business in the banking sector.

Some authors have extended the model to the "5V" concept, which includes two additional characteristics of data: Value and veracity. Value means that it is not enough to have a large amount of data, but that this data must also have some value, i.e. its disaggregation and analysis should create a reasonable benefit for data users. Otherwise, such an amount of data could be considered useless. In the financial industry, this means deep segmentation in the core banking system in certain areas, e.g. loan applications should be differentiated between retail and corporate portfolios, then within the retail area between households and other retail customers, while within the corporate area a distinction is made between micro-enterprises, small and medium-sized enterprises and large enterprises and so on.

Veracity refers to the quality or reliability of the data itself [5]. Reliability in the banking sector means that banks should be able to check the quality of the data collected from external sources and, based on the assessment made, incorporate the filtered data into the core banking system with a satisfactory level of reliability and appropriate values.

In some papers, authors emphasise the existence of the "7V" concept for Big Data and add the components of variability (related to the inconsistency of data) and visibility (the process of presenting data in a visually acceptable format, such as charts and tables) [6]. To make their workflows more efficient, banks should scrutinise the consistency level of the data collected. Only filtered data should be fed into the bank's core banking system. After all structured, unstructured or semi-structured data in the core banking systems has been transformed into useful data, this data is available for various processing options to create graphs, charts or other types of reports that fulfil the seventh V in the "7V" concept, visibility.

The described characteristics of the Big Data concept and their implementation in the banking sector could be an important factor for competitive advantages for banks. The practise of using Big Data in banks is the focus of the following section.

3. BIG DATA AND ADVANCED ANALYTICS IN BANKING INDUSTRY

According to European Banking Authority, which published report on Big Data and Advanced Analytics in banking industry, there are four key pillars as follows [7]:

- Data management (type, source, protection, and quality of data),
- Technological infrastructure (data platforms, infrastructure, and processing),
- Organization and governance, and
- Analytics methodology (development, implementation, and adoption of Advanced Analytics solutions).

Digital transformation in banking industry and its dependence upon Big Data is the subject of various analyses especially in dramatically changing business world. Four key pillars in Big Data are represented in Figure 1.

Data management is about controlling and securing data within an organisation. It includes different types of data (from structured to semi-structured to unstructured) as well as data that comes from internal or external sources. Banks are often focused on the collection and use of internal data, e.g. data on customer transactions, data on credit card usage, data on loan repayment behaviour, etc.

Figure 1. Key pillars in Big Data and Advanced Analytics. [7, p. 7].

External data, on the other hand, refers to financial data for business registers, demographic data from statistical offices, data from credit bureaus and so on. Before final use, it is important to validate the external data collected. Data protection and data security are also very important. Therefore, appropriate security precautions, including organisational and technical measures, should be taken. Finally, data quality should be considered in all financial institutions, especially in the data collection and preparation phase. Data that comes from dubious sources could lead to serious errors and inappropriate results, resulting in inappropriate decision-making.

The second pillar relates to three components: Infrastructure, Data Platform and Processing. It is important to emphasise that in the processing phase, volume and velocity are significantly represented as "2Vs" of the "7V" concept to support Advanced Analytics. The third pillar is related to the introduction and implementation of a governance structure and an appropriate organisation through the definition of solid roles and responsibilities while respecting transparency principles. Continuous training for managers and well-defined communication channels with the management could increase the efficiency of the banking organisation. Having recognised the importance of the development of Big Data and Advanced Analytics in banks, management could be proactive and encourage employees to share their knowledge and improve all aspects of the workplace where Big Data has a major impact.

The fourth pillar is analytics methodology which involves: development, implementation and adoption of Big Data and Advanced Analytics. As per [7] there should be encompassed four main phases: data collection, data preparation, analytics, and operations, which are presented in Figure 2.

Advanced Analytics consider usage of machine learning and its algorithms in processing of large amount of data. Abovementioned is illustrated in Figure 3.

Data can be collected from various sources, while the raw data is transformed into a form that makes it available and ready for further analysis during the data preparation phase. In the analysis phase, ML techniques are used to develop models that extract knowledge from the data. Finally, in the operational phase, the end user can gain insight into the results of the model and monitor them over time in order to always obtain accurate results. It is necessary to accurately describe the aspect of how ML algorithms work. ML algorithms can be categorised into four main groups from the perspective of human involvement in data labelling as follows [8]:

1. supervised learning,
2. unsupervised learning,
3. reinforcement learning, and
4. deep learning.

Supervised learning assumes that the data set contains labels for certain observations that provide the algorithm with the general rule on which the classification of the data is based.

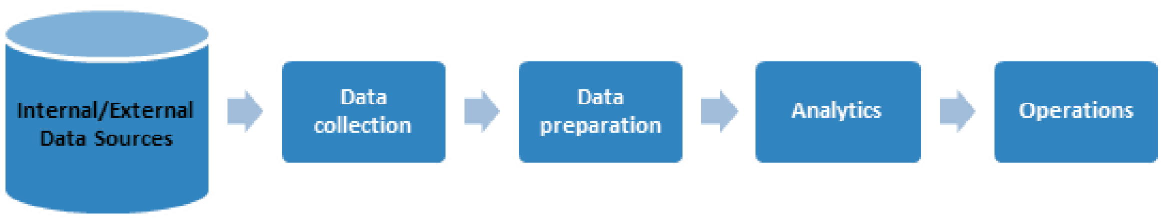

Figure 2. Advanced Analytics methodology. [7, p. 5].

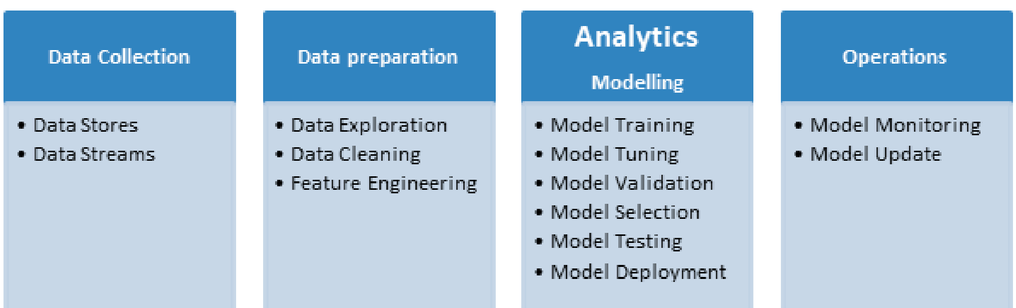

Figure 3. Advanced Analytics process using machine learning. [7, p. 31].

This rule is then used to predict results and recognise patterns for the rest of the observations. An example of supervised learning is spam detection in the enterprise, as it is based on "training" databases to deal with new data, which should be categorised into spam and non-spam related data. Unsupervised learning means that the dataset does not contain any labels, so the algorithm should only find the pattern (rule) by forming clusters that have similar characteristics. Examples of unsupervised learning include customer segmentation for the purpose of targeting when implementing marketing strategies and the classification of users based on their activities on social media. Reinforcement learning is a mixture of monitored and supervised learning. It starts with an unlabelled data set, whereupon the algorithm selects an action for each data point and receives feedback ("learning") from the human. The use of automated robots in certain areas is an example of reinforcement learning, e.g. the use of robots in restaurants that bring food to the tables. Deep learning is an ML technique that teaches computers to learn by example. Deep learning consists of a neural network with three or more layers: Input layers, hidden layers and output layers. Deep learning algorithms can be used for supervised, unsupervised or reinforcement learning. The results of deep learning are evident in image recognition and natural language processing [8].

ML algorithms in the financial industry could be used to fulfil regulatory requirements in a more modern way by applying cost-efficient strategies in different banking areas. They could also be used in the "front office" (e.g. insurance, credit scoring) or in the "back office" of certain banks (e.g. market impact analysis, risk management modelling) or in trading and portfolio management [8]. The next section looks at examples of the use of ML in the banking sector, focussing on the automation of decision-making processes in specific banking areas.

4. IMPLEMENTATION OF ML IN BANKING INDUSTRY

Banks, as financial intermediaries between companies with cash surpluses and deficits, are creating their market position in the modern era by adapting their business to the changes required by the new Big Data era. All functions within banking organisations are undergoing a serious transformation in the world of Big Data in terms of technology and the use of Advanced Analytics as the main driver of future risk management in banks.

In contemporary environment, banks have access to large amounts of customer data that can significantly improve the decision-making process. Access to external, unstructured data is the basis for better decisions in assessing customers' creditworthiness and credit risk, but also the basis for monitoring the quality of bank portfolios and developing early warning systems, identifying financial crime and predicting operating losses. The rapid adoption of various models used in large-scale data analysis provides a more detailed view of all available data, making the model itself more accurate and precise.

ML techniques have been widely used in the banking sector and by financial institutions in general. The most widespread areas for the use of ML are related to fraud detection, anti-money laundering and real-time payment monitoring. In the area of credit risk, the application of ML techniques usually means assessing the creditworthiness of customers or determining the credit scoring for customers at loan disbursement. More specifically, in the area of risk management to determine capital requirements for credit risk, banks use various types of regression analyses or decision trees and other ML techniques when implementing the IRB. The application of ML techniques could be very useful for credit approval procedures as these techniques help to improve the predictive power of IRB models. Certain financial institutions use ML techniques for model evaluation, collateral valuation or probability to default (PD) estimation. The most obvious challenges in the efficient application of ML are usually the quoted results and their interpretation, their appropriate transmission and reporting to management and finally the justification of the results to regulators [9].

ML has undoubtedly become a major force in the banking sector, driving improvements in decision-making processes. Instead of decision making based solely on the human factor and personal involvement, ML has contributed to a more sophisticated yet objective and consistent system of decision making. This system could include various techniques such as decision trees or neural networks, which have a large amount of data that provides an excellent basis for creating predictive models. In addition to the positive effects of ML, which are mainly reflected in the more accurate and precise work with Big Data and the improvement of efficiency in all areas of banking, it should be noted that ML techniques also have their weaknesses. The main disadvantages of using ML in the banking sector include: overfitting, lack of explainability and potential bias [10]. Overfitting of the model occurs when a large amount of data is entered into the model, so that too many details have a negative influence and the model does not work well due to the complexity of the decision trees. The model must be explainable, i.e. people must be able to understand how it arrives at a certain result and what the basis for this result is. Conversely, the model could not be justified in a simple way and it could not be interpreted appropriately. Potential bias is related to avoiding discrimination in the processing of data (based on age, gender, race or religion) and achieving fairness (non-discrimination).

Aside from the shortcomings, the contribution of Big Data and Advanced Analytics in the areas of fraud detection, anti-money laundering and, in particular, risk management is negligible. Identifying and using decision trees based on pre-defined rules and minimising human involvement is evident in the validation of PD models, collateral valuation and regulatory compliance when banks use the IRB model.

5. CONCLUSION

The use of modern technologies and Advanced Analytics are the main drivers of future risk management in banks. Working with Big Data represents a new challenge for banks in many respects. Access to unstructured data could be an excellent basis for better decision-making processes and the assessment of customers' creditworthiness and credit risk. By validating raw data after the collection and processing phase and analysing it with ML techniques, the entire process of doing business in banks becomes more efficient and effective. In addition, the advancement of artificial intelligence and the application of ML techniques have facilitated the determination of capital requirements for credit risk and enabled a more efficient performance of model assessment, collateral valuation and PD estimation, contributing to a better fit of IRB models. Of course, adequate support from ML is not enough if the results obtained cannot be passed on to management in the right way.

As most risks in banking management nowadays take it origin from credit risk, future research will be dedicated to challenges brought by the application of the Big Data concept emphasizing analysis of counterparty credit risk, especially usage of ML algorithms to calculate the default on derivative transactions.

6. REFERENCES

[1] J. Lukić Nikolić and V. Mirković, “Transformacija poslovnih procesa i upravljanja bankarskim rizicima u Big Data svetu,” [“Trasformation of business processes and risk management in banks in Big Data world,”] in XXIII Internacionalni naučni skup SM 2018, Strategijski menadžment i sistemi podrške odlučivanju u strategijskom menadžmentu, Subotica: Ekonomski fakultet, pp. 137-144., 2018.

[2] V. Mirković, J. Lukić Nikolić and V. Martin, “Reshaping Banking Industry Through Digital Transformation,” in FINIZ 2019 - Digitization and Smart Financial Reporting, Belgrade, Singidunum University, Serbia, 2019, pp. 31-36., doi: 10.15308/finiz-2019-31-36.

[3] V. Mirković, “Uticaj finansijske tehnologije na savremeno bankarsko poslovanje,” [“The impact of financial technology on modern banking business,”] in XXII Internacionalni naučni skup SM 2017, Strategijski menadžment i sistemi podrške odlučivanju u strategijskom menadžmentu, Subotica: Ekonomski fakultet, pp. 304-311., 2017.

[4] M. Minelli, M. Chambers and A. Dhiraj, *Big Data, Big Analytics*. New Jersey: John Wiley & Sons, 2013.

[5] S. Gutta, ”Data Science: The 5V's of Big Data,” 04 05 2020[Online]. Available: https://medium.com/analytics-vidhya/the-5-vs-of-big-data-2758bfcc51d [Accessed: 21 01 2024]

[6] V. Niculescu, “On the Impact of High Performance Computing in Big DataAnalytics for Medicine,” *Applied Medical Informatics*, Vol. 42, No. 1/2020, pp. 9-18, 2020.

[7] European Banking Authority, “EBA Report on Big Data na Advanced Analytics,” January 2020 [Online]. Available: https://www.eba.europa.eu/sites/default/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf [Accessed: 24 01 2024]

[8] Financial Stability Board, “Artificial intelligence and machine learning in financial services,” 01 11 2017 [Online]. Available: https://www.fsb.org/wp-content/uploads/P011117.pdf [Accessed: 24 01 2024]

[9] European Banking Authority, “Machine Learning for IRB models,” August 2023 [Online]. Available: https://www.eba.europa.eu/sites/default/files/document_library/Publications/Reports/2023/1061483/Follow-up%20report%20on%20machine%20learning%20for%20IRB%20models.pdf [Accessed: 24 01 2024]

[10] R. Fairushyn, “Application of machine learning methods in the banking sector to increase the efficiency of decision-making,” *Telecommunication and information technologies,* 79(2), pp. 76-85, DOI: 10.31673/2412-4338.2022.027685, 2023.