Oct 26th, 1:30 PM - 3:00 PM

An Overview of Big Data Analytics in Banking: Approaches, Challenges and Issues

Fisnik Doko
Univiversity Ss. Cyril and Methodius, fisnik.doko@gmail.com

Igor Miskovski
Univiversity Ss. Cyril and Methodius, igor.miskovski@finki.ukim.mk

An Overview of Big Data Analytics in Banking: Approaches, Challenges and Issues

Fisnik Doko[1], Igor Miskovski[1], Miroslav Mirchev[1]

[1] Univ. Ss. Cyril and Methodius, Faculty of Computer Science and Engineering, Skopje

Abstract. Banks are harnessing the power of Big Data. They use Big Data and Data Science to drive change towards data and analytics to gain an overall competitive advantage. The Big Data has potential to transform enterprise operations and processes especially in the banking sector, because they have huge amount of transaction data. The goal of this paper is to give an overview of different approaches and challenges that exists in Big Data in banking sector. The work presented here will fulfill the gap of research papers in the last five years, with focus on Big Data in central banks and credit scoring in central banks. For this paper, we have reviewed existing research literature, official reports, surveys and seminars of central banks, all these related directly or indirectly to Big Data in banks.

Keywords: Big Data, Banking, Credit scoring

Introduction

There is no single official definition for Big Data, it is referred as collection and analyses of large volumes of structured and unstructured data, potentially in real time to create value for companies. Concept of Big Data in financial context is different from other industries [l]. Popular approach for the pillars of Big Data are the Gartner 3V model, and the IBM infographicl that provides an overview of the components of "Big Data" by adding a fourth "V". Big Data is mostly defined with the following properties known as "five Vs" [2], [3]:

1. Volume: in today's connected world, huge amounts of data is being created every second, from tweets to videos and photos to bank transactions. Financial services [4] during the years handle high volume of data and always have been with the biggest datasets.
2. Variety: the data can range from structured to unstructured. Big Data technology allows users to analyze not only the structured data we find in the financial sector, but also the more complex unstructured data that is becoming more relevant in order to discover new conclusions and findings.
3. Veracity: refers to the trustworthiness of the data in Big Data, especially if it is obtained from third-party public sources. Veracity burden can rise exponentially with data volumes.
4. Velocity: is a concept which deals with the speed of the incoming data from different sources, it is the data is frequently updated and can be quickly analyzed with possibility to real time analyzes.
5. Value: by predicting new trends based on the analysis of data, banks and financial services can create value for customers by offering them new services.

Big Data is the modern emerging field where technology and Data Science provide new ways of extracting value from the ocean of new information. The ability to effectively manage

insights and extract knowledge is a key competitive advantage [5]. Finance industry uses Data Science for supporting and predicting credit scoring and trading and in operations via fraud detection [6].
Banks need to leverage the benefits of the volume of data they collect and the digital revolution trends to provide better adapted services to their customers in an increasingly competitive digital world [7].
In Section 2 we overview some of the usages and benefits of Big Data in banking industry. In Section 3 we overview the major challenges in the banking industry for adopting Big Data. Then, in Section 4, we describe the usage of Big Data in central banks. After that, in, in Section 5, we provide related work for credit scoring in the last years. In Section 6 we summarize the conclusions.

Advantages of Big Data in banking

Big Data and Data Science in banking refers to the field of applying software and technology in combination with advanced algorithms and methods to gain more insights, to make informed decisions, or to predict risk and revenue. Initial efforts are focused on gaining new insights from existing data and then by newly available sources of information with top priority customer centric objectives [4].
The following shows a list of advantages of Big Data in banking process and Data Science use cases that have the highest impact on the banking sector harnessing benefits of these technologies in 2018-2019 [8] [9]:
• Risk management. Big Data can be targeted to an organization's needs and applied to enhance different risk domains: credit risk, liquidity risk, operational risk and market risk [10].
• Customer insights, experience and analytics. Banks are looking to leverage large amounts of customer data across multiple service delivery channels to discover customer behavior patterns and increase knowledge for consumers [11].
• Enhanced fraud detection. Big Data technologies enable real-time analytics on larger datasets and correlation of data from multiple sources to determine fraud more efficiently [12].
• Algorithmic trading and forecasting stock market. Combining of various data sets from multiple markets and geographies provides enhanced view of market which can generate trade signals, trade execution, profit and risk exposure.
• Predictive analytics. Reveals patterns in the data that foresee the future event that can happen, through understanding social media, news trends, and other data sources, predicting prices and customers lifetime value and the market moves.
Following are Big Data use cases for marketing in the banking industry [13]:
• Sentiment analytics. Monitor social media to increase marketing success and to adjust marketing tactics correctly, identify high influence customers in social media because they are critical to fulfil goals [14].
• Customer 360. Identify the customer profile using more attributes to investigate the habits and to build complete holistic customer profile. Understanding the product engagement of the customer to send the correct marketing message.
• Customer segmentation. Big Data enables faster and sharper classification of customers into various segments that share similar characteristics or behaviors
based on consumer behavior and different attributes [15].
• Next best offer. Allows an organization to increase its opportunities by predicting what the customer wants next using recommendation system to predict customer preferences (like Netflix) based on linking of historical transactions to products.

Challenges and issues

Adopting transformational potential of Big Data is very complex process which implies many changes in IT ecosystems of banks. There are numerous challenges can face during this integration such as infrastructure, information privacy and storage cost. Some generalized challenges that should be overcome to have successful Big Data project are given in the following text.

Privacy and security. Handling Big Data is more scalable and flexible in cloud, but the privacy and security regulations often restrict this movement decision, also there are reputation consequences. Big Data has faced criticism for overstepping privacy boundaries. Ensuring Big Data projects retain their integrity and trust is critical to avoiding public embarrassment, mistrust, and liability. Big Data analytics is limited with numerous regulatory compliance considerations for data protection and privacy issues which impact the analyses because of the individuals being able to decline banks to use their personal data from processing in certain circumstances. Ethical issues should be reviewed in such scenarios, because predictive analyses will identify people with low social conditions which can be treated worse [16].

Storage and processing issues. While banking structured data are continually growing, the unstructured data is growing faster and is becoming more important source for customer insights [4]. This increases the need for having unstructured terabyte databases. The conventional data management techniques are no longer enough to handle the massively large, high-velocity, heterogeneous financial dataset [17]. Uploading large data in cloud or using cloud data for real time analyses with data that reside on-premise, it has processing performance impact that may not be the right choice. The biggest issues when using cloud are privacy and regulatory implications and regulations.

Technical and architectural challenges. Data is rapidly increasing hence it is very important to use appropriate techniques and technology that can handle such vast amount of large and variety complex datasets, which requires new infrastructure components like Hadoop, NoSQL, Map Reduce, where banking industry is lagging behind their peers in other industries [18]. Reading/writing data in enterprise SAN storage and replicating to redundant copies is wasteful for Big Data. These challenges
and the chances for failure are easily handled by the Hadoop Distributed File System
and by the Map Reduce programing paradigm. High-level development environments that abstract from processing architecture, are programming languages such as R and Python which are used to discover new insights from data [19].

Analytical challenges. Big Data itself does not create value until analytical capabilities are used that require skills to use them. Banks need skilled staff data scientists to benefit from Big Data opportunities and overcome governance issues. Because of strict governance rules in banking, banks are missing special working positions for data scientists. Also because of their standards and regulations banks are behind the other technologies in analytical capabilities of text and sentiment analyses,
voice and video. Also, the lack of multiple data types is factor where banks lag behind other industries.

Big Data in central banks

Big Data in central banks is of high volume because data are reported in granular basis on transaction level for example credit data for person or security by security, velocity is argued with the frequent gathering of these data. Drawback is that central banks have more general aggregated data and the data sets are often just numerical.

Central banks already have large datasets of statistics, structured data and information, which are regularly used into their decision-making process. Credit registry is often the largest data sets maintained by some central banks.
Experience of Central banks with Big Data is surveyed from BIS-IFC Big Data Survey [20]. As reported in 2015 there wasn't clear understanding of the "Big Data" definition. The primary focus to central banks has been has been accessing and processing the data [20].
One promising survey conducted by Central Banking in association with BearingPoint during mid-2017. It proves that Big Data is an active area for new projects in central banks and key Big Data projects in central banks is reported development of credit registers, followed by administrative sources and consolidation of internal systems [21]. In the IFC annual report 2018 [22] the feedback from central banks is concerned with the complex privacy implications of dealing with Big Data. The last
survey conducted in 2018 by Central Banking in association with BearingPoint [23]
reports on the approach central banks take towards Big Data. Central banks started looking for external sources to obtain Big Data. Complex privacy implications of dealing with Big Data and other sources as privacy laws evolve with the time and Big Data in central banks is predominantly regarded as useful for research and there is no evidence involvement in policymaking.

Credit scoring

Credit risk is the probability of loss due to a borrower's failure to make payments on any type of debt. The survey of Economist Intelligence Unit shows that most of the banks surveyed in six continents, especially Europe, are primarily concerned with the credit risk [10]. Models can be improved when there are other data sources with the ability to combine and join data from multiple sources. A well-trained system can then perform the credit-scoring tasks and help employees work much faster and more accurately.
Paper [24] shows that the neural network-based credit scoring model is more effective in screening default applications. Paper [25] introduces a two-step loan credibility prediction system which uses Decision Tree Induction algorithm for prediction. Paper [26] presents a new combination approach based on consensus of six classifier that creates a group ranking. Paper [27] shows that the usage of client's payment data is very important factor to enhance credit score prediction. Paper [28] demonstrates that most significant attributes in determining the outcome of a credit application are income, years of employment, credit score and whether or not the applicant had defaulted on a prior credit account. Paper [29] shows that data from social media is big factor in determination of credit score. The newest papers [30] and [31] presents greater insight on the usage of Big Data for credit scoring models, and how this technology is overcoming traditional credit scoring models. This provides proven base that banks should develop new credit models and introduce new data sources that will significantly enhance the model. They propose Big Data models which will include various data for the clients including where they shop, their purchases, their online social network profiles and other factors that aren't directly related with credit default.

Conclusion

Banks will realize value by effectively managing and analyzing the rapidly increasing data and putting the right skills and tools to better understand their operations and customers. By harnessing Data Science tools and technologies, banks can more effectively inform strategic

decision-making, reducing uncertainty and eliminating analysis-paralysis. Despite all the advantages of Big Data, it has a set of limitations when it comes to implementation in banking. Retail and central banks are using Big Data in a variety of ways that they treat as a Big Data, but they have to overcome all the mentioned challenges to be in hop with technology and to gain maximum benefit from Big Data. However, banks are still lagging in implementing Big-data credit- scoring tools will emerge as a way to ensure greater efficiency in underwriting while expanding access to the underbanked and to historically selected groups. Big Data is the reality and is going to stay there for a long time. Banks need to continuously revise policy and regulatory standards to can adopt new technologies. The above analysis shows that there are still research challenges to develop at all levels of the Big Data chain and involve a wide set of different technologies. Despite the technological aspects, there are organizational, cultural, and legal factors that dictate how banks will continue the road of implementing Big Data operations and business development

References

1. B. Fag and P. Zhang, "Big data in finance," in In Big data concepts, theories, and applications, Springer, Cham, 2016, pp. 391-412.
2. I. Kalbandi and J. Anuradha, "A brief introduction on Big Data 5Vs characteristics and Hadoop technology," Procedia Computer Science 48, pp. 319-324, 2018.
3. S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends [point of view].," Proceedings of the IEEE, vol. 103.2, pp. 143-146, 2015.
4. I. I. f. B. Value, "Analytics: The real-world use of big data in financial services," IBM, 2016.
5. S. Dubey and S. Nainwani, "https://www.cognizant.com/InsightsWhitepapers/ Banking-on-Data-Science.pdf," Cognizant, 2019. [Online].
6. G. E. F. D.-M. a. J. D. A.-L. Melo-Acosta, "Fraud detection in big data using supervised and semi-supervised learning techniques.," IEEE Colombian Conference on Communications and Computing (COLCOM), no. IEEE, 2017.
7. S. Totman, "Fintech Business," July 2017. [Online]. Available: https://www.fintechbusiness.com/blogs/759-big-data-and-customer-engagement. [Accessed 2019].
8. Srikanth, "Techiexpert," June 2018. [Online]. Available: https://www.techiexpert.com/top-data-science-use-cases-in-finance/. [Accessed May 2019].
9. E. Moldavskaya, "Intetics," Intetics Inc, November 2018. [Online]. Available: https://intetics.com/blog/top-5-machine-learning-use-cases-for-the-financial-industry. [Accessed May 2019].
10. "Retail banks and big data: Big data as the key to better risk management," The Economist Intelligence Unit, 2014.
11. K. Egetoft, "Digitalist Magazine," January 2019. [Online]. Available: https://www.digitalistmag.com/customer-experience/2019/01/29/data-driven- analytics-practical-use-cases-for-financial-services-06195123. [Accessed May 2019].
12. L. S V S S and S. D. Kavila, "Machine Learning For Credit Card Fraud Detection System.," International Journal of Applied Engineering Research, vol. 13, pp. 16819-16824, 2018.

13. Y.-P. SHEE, D. CROMPTON, R. HILLE and S. PETTER MÆHLE, "EVRY," [Online]. Available: https://www.evry.com/globalassets/insight/bank2020/bank-2020- --big-data--- whitepaper.pdf. [Accessed May 2019].
14. M. Pejić Bach, Ž. Krstić, S. Seljan and L. Turulja, "Text Mining for Big Data Analysis in Financial Sector," Sustainability, vol. 11, 2019.
15. "Medium," ActiveWizards, May 2018. [Online]. Available: https://medium.com/activewizards-machine-learning-company/top-9-data-science- use-cases-in-banking-6bb071f9470c. [Accessed May 2019].
16. A. Chandani and M. Mehta, "BANKING ON BIG DATA: A CASE STUDY," ARPN Journal of Engineering and Applied Sciences, vol. 10, 2015.
17. D. Bholat, "Big Data and central banks," Big Data & Society, 2015.
18. A. Jaiswal and P. Bagale, "A Survey on Big Data in Financial Sector," International Conference on Networking and Network Applications (NaNA), no. IEEE, 2017.
19. R. Elshawi , S. Sakr, D. Talia and P. Trunfio, "Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service.," Big Data Research., 2018.
20. "Central Banks use of and interest in big data," Irving Fisher Committee on Central Bank Statistics, October 2015. [Online]. Available: http://www.bis.org/ifc/publ/ifc- report-bigdata.pdf. [Accessed May 2019].
21. "Big data in central banks: 2017 survey," Central Banking, October 2017. [Online]. Available: prod.centralbanking.bb8.incinsight.net/central-banks/economics/data/3315546/big-data-in-central-banks-2017-survey. [Accessed May 2019].
22. "2018 IFC Annual Report," Irving Fisher Committee on Central Bank Statistics, 2018. [Online]. Available: https://www.bis.org/ifc/publ/ifc_ar2018.pdf. [Accessed May 2019].
23. E. Glass, "Central Banking," Big data in central banks: 2018 survey results, August 2018. [Online]. Available: https://www.centralbanking.com/central-banks/economics/data/3661931/big-data-in-central-banks-2018-survey-results. [Accessed may 2019].
24. A. Byanjankar, M. Heikkilä and J. Mezei, "Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach," IEEE Symposium Series on Computational Intelligence, pp. 719-725, 2015.
25. S. M, "Two step credit risk assessment model for retail bank loan appli-cations using decision tree data mining techniques," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 5, no. 3, pp. 705-718, 2016.
26. M. Ala'raj and M. F. Abbod, "Classifiers consensus system approach for credit scoring," Knowledge-Based Systems, vol. 104, pp. 89-105, 2016.
27. E. Tobback and D. Martens, "Retail credit scoring using fine-grained payment data.," Journal of the Royal Statistical Society Series A (Statistics in Society), 2019.
28. A. Smith, D. Khaneja and B. Maher, "Credit Approval Analysis using R," 2017.
29. J. Lohokare, R. Dani and S. Sontakke, "Automated Data Collection for Credit Score Calculation based on Financial Transactions and Social Media," International Conference on Emerging Trends & Innovation in ICT (ICEI), pp. 134-138, 2017.
30. "A review of credit scoring research in the age of Big Data," JOURNAL OF FINANCIAL REGULATION AND COMPLIANCE, vol. 26, no. 3, 2018.
31. M. Hurley and J. Adebayo, "Credit scoring in the era of Big Data," Yale Journal of Law and Technology, vol. 18, pp. 148-201, 2017.

32. J. M. Cavanillas, E. Curry and W. Wahlster, New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe., Springer, 2016.