Approach to the choice of Big Data processing methods in financial sector companies

Igor Ilin^{1*}, Oksana Iliashenko¹ and Victoria Iliashenko¹

¹Peter the Great St.Petersburg Polytechnic University, Polytechnicheskaya, 29, St. Petersburg, 195251, Russia

Abstract. The paper describes the tasks of financial sector companies, solved using the technology of processing Big Data. The possibilities of using Big Data for solving the problems considered are analyzed. A comparative analysis of the methods for processing Big Data has been carried out, a correspondence has been established between the methods considered and the tasks being solved by the companies of the financial sector. Criteria for choosing methods for processing Big Data are developed depending on the type of tasks, instrumental support of the proposed approach has been performed.

1 Introduction

The main tasks of modern companies in the financial sector, and above all of the banking institutions, are to achieve a new level of competitiveness that allows them to compete with technology companies, while remaining in demand in terms of providing services to customers. A key factor in the successful implementation of digital transformation financial sector companies is the construction of an ecosystem in which it is necessary to create a universal technological platform. Such a platform aggregates the producers of goods and services around itself and, on the basis of a behavioral characteristics analysis of customers, forms the best offer for them. In this regard, the main tasks of the financial sector today are prompt reporting, scoring, fraud prevention, personalization of banking products, obtaining information about customer behavior, their habits, welfare and consumer behavior. Successful solution of these tasks involves the timely processing of Big Data. Successful solution of these tasks involves the timely Big Data processing. Financial institutions generate a huge amount of data daily, such as purchase history, profile data, browsing history or social networking data. In this regard, there are questions of studying and selecting the necessary software and suitable methods for Big Data processing in accordance with the requirements of the legislation of the RF and the Central Bank of the Russian Federation.

With reasonable management, these data can be used for further analysis, which will effectively achieve its main marketing goals [1, 2].

The massive introduction of Big Data analysis technologies is complicated by the fact that financial sector companies often use disparate or obsolete platforms. The relevance of the study is due to the modern development of software and hardware technologies that create

* Corresponding author: olgakalinina@bk.ru

the need to organize work with huge amounts of data relating to external and internal processes of organizations. This leads to the creation of a huge data flow, passing through the servers of financial companies, which carries a certain set of useful data. With the correct processing of data, they are transformed into information that can be an advantage in the financial services market. Large companies are forced to organize entire departments to study and select the necessary software and suitable processing methods in accordance with the requirements of the Central Bank of the Russian Federation, and the requirements of the Bank's Business. The content and development of such departments is the main part of the company's expenses [3, 4].

2 Technologies, methods and tasks

Consider the existing technologies and methods for processing and analyzing large data that are sold in financial sector companies.

2.1 Technologies

Technologies used in the Big Data processing can be divided into several groups: software, process equipment and services.

The software group is the most common for processing data, it includes the following approaches:

1. A scripting language for statistical data processing - R, which allows you to work with data, with help of which you create and modify data for analysis.

2. NoSQL consists of a number of approaches that implement databases that have distinctive features from fate and relational DBMSs.

3. MapReduce is a computational model that allows you to compute very large sets of data in parallel.

4. Hadoop has a great advantage, because each block was copied more than once, which protects the system from failure, with its help, search and contextual mechanisms of highly loaded sites are implemented.

5. SAP HANA - the NewSQL platform, which is distinguished by its performance for storing and processing data [5,6].

The group of technological equipment includes servers that include data warehouses and infrastructure equipment that allows to speed up the platform, provide constant power, including server console kits and similar equipment for the physical layer. The last group is services. They help with the arrangement and optimization of the infrastructure, and also provide data protection. Big Data processing technology also includes processing and analysis methods that can be combined in one tool or at the same level of architecture, for example, at the filtering level.

2.2 Methods

The main methods of processing and analysis include the following methods and techniques::

1. Statistical analysis is a method of detecting interesting correlations between variables in Bid Data. It was first used by large supermarket chains to discover interesting relationships between products using data from supermarket outlets. Typical problems solved by statistical analysis:

- place products in the best proximity to each other to increase sales;
- retrieve information about visitors on websites from the logs of the web server;
- analyze biological data to identify new relationships;

- monitor system logs to detect intruders;
- determine whether people buying milk and butter can buy diapers.

2. Machine learning - the creation of algorithms for self-learning, based on the data received. Machine learning includes software that can learn from data. This allows computers to learn without being explicitly programmed. Machine learning is focused on making predictions based on known properties derived from sets of "data for learning". Most often used in the following tasks:

- find differences between spam and regular e-mails;
- study user settings and make recommendations based on the information received;
- identify the best content for attracting potential customers;
- determine the probability of winning [7].

3. Data Mining is a collective name used to denote a set of methods for detecting previously unknown, non-trivial, practically useful and accessible interpretation of knowledge necessary for decision-making in various spheres of human activity in data. Used to solve problems through data analysis. Data Mining Tools enable enterprises to predict future trend:

- to create risk models and detect fraud;
- to improve product safety, identify quality problems, manage supply chains and improve operations [8, 9].

Technologies and methods Big Data processing are closely interrelated with the tools that use them. Some tools for analyzing Big Data include several methods for analyzing large data.

2.3 Tasks

Big Data in the financial sector offer a wide range of solutions for all areas of the company:

- company's reputation analysis;
- customer-focusedness;
- compliance with the Central Bank requirements;
- modernization of basic banking systems;
- increased efficiency;
- management of risks.

The financial sphere generates a huge amount of data, where each interaction with the customer generates electronic records that must be stored because of regulatory requirements. Thanks to the use Big Data analyzing methods, data is simply not stored as needed, but is actively used to generate business ideas that enhance the value of the company. Data analysis now occurs in real time, which affects the situation on time. Let's consider the typical five tasks on the use of Big Data in the financial sphere [10].

1. Fraud detection.

Use of Big Data to Differentiate Fraudulent Interactions from Legal Transactions. Analysis systems involve immediate actions, such as blocking irregular transactions that stop fraudulent transactions before it ends and not allow not lose customers and profits because of such transactions.

2. Requirements for regulations compliance.

The financial sector companies operate in a tight regulatory environment, which requires high-performance levels of monitoring and reporting. The Dodd-Frank law, adopted after the financial crisis of 2008, requires monitoring transactions and documenting the details of each trade. These data are used to monitor trade, in which abnormal trading models are recognized.

3. Customer segmentation.

In order for financial sector companies to remain competitive, they move from product-oriented customers to customer-oriented products. One way to achieve this transformation is to understand your customers through segmentation. Big Data allows you to group customers into separate segments that are defined by data sets that can include customer demographics, daily transactions, interaction with online customer service systems and telephones, and external data such as the cost of their housing. Then, orient the advertising and marketing campaigns to customers in accordance with their segments.

4. Custom offers.

One of the steps beyond segment marketing is personalized marketing, aimed at customers, based on an understanding of their individual buying habits. Although it is supported by a large analysis of trade record data, financial services firms can also include unstructured data from their customer's social networking profiles to create a more complete picture of customer needs by analyzing customer sentiment. After understanding these needs, a Big Data analysis can create a credit risk assessment to decide whether to continue the transaction. For example, Nedbank Ltd is a large bank in South Africa, which realizes its advantages using analytics in social networks. Analysis of various platforms in social networks in almost real-time mode provides Nedbank Marketing Department with information about the marketing campaign, preferences and complaints of customers.

This technology significantly reduced the cost of monitoring social networks while increasing marketing success.

5. Management of risks.

Despite the fact that each sector of the economy must deal with risk management, the greatest need for the financial industry. Schemes and regulatory documents, such as Basel III, require firms to manage liquidity risk in the market through stress testing. Financial companies also manage their risk to customers through the analysis of complete client portfolios.

The risks of algorithmic trading are controlled through testing strategies against the analysis of historical data. Analyzing Big Data will help to generate criteria and give an alert in real time mode if the risk threshold for the organization is surpassed.

Now there are more than 25% of financial companies that have implemented the Big Data, so they gain a competitive advantage. Due to both regulatory requirements and the perceived value of Big Data, the financial sector will continue to implement huge projects to analyze large amounts of data. This will require increased investment in data center technology, as well as increased demand for personnel with the skills to work with Big Data.

3 Results

The main tasks for which banks use Big Data analysis technologies are prompt reporting, scoring, avoidance of questionable transactions, fraud and money laundering, as well as personalization of banking products offered to customers.

Consider the tasks in Table 1, for which banks have chosen a way of processing and storing data using Big Data. It can be seen from the table that each bank uses Big Data for various tasks [11].

For convenience, the analysis of the links between processing methods and tasks can be broken down according to the list of typical tasks.

Table 1. Tasks solved in banks with the help of Big Data.

	Name	Tasks		
1	Sberbank	Fight fraud	Management of risks	Segmentation and customer credit score
2	Alfa-Bank	Analysis of social networks and the behavior of users of the site	Customer Credit Score	Forecasting of customers outflow
3	VTB Bank	Analysis of customer feedback in social networks and forums	Formation of financial statements	Segmentation and customer outflow management
4	Gazprombank	Scoring	Fraud counteraction	Personalization of offers
5	Tinkoff Bank	Scoring	Management of risks	Marketing programs

Enlargement in this case is applicable, because the method of processing a common task can not differ from its subtask. The relationship between typical tasks and the tasks listed in Table 1 is shown in Table 2.

Table 2. Correspondence of the realized tasks and sub-tasks to typical tasks

Typical tasks	Differentiated tasks
Fraud detection	Fraud counteraction
	Scoring verification of potential borrowers
Requirements for regulations compliance	Formation of financial statements
	Optimization of internal processes
	Forming daily reporting
	Rapid reporting
Customer segmentation	Segmentation and customer outflow management
	Analysis of customer feedback in social networks and forums
	Analysis of social networks and the behavior of the site users
	Analysis of marketing companies results
	Scoring
	Improving the quality of customer service
Custom offers	Personalization of offers
	Marketing programs
	Customer Credit Score
	Create personalized loan offers
Customer segmentation	Forecasting of customers outflow
	Sales process optimization
	Cost optimization
	Forecasting queues in offices

After integration of tasks, it became possible to evaluate what data processing methods they can be analyzed and processed. To this end, consider the methods for processing Big

Data described earlier in this paper and compare them with typical tasks that need to be solved with their help.

Table 3 shows the use of processing methods. Big Data applied to typical tasks based on relevance of use, this means that most methods are applicable to all tasks, but they may not always produce correct results or they may not be suitable for the type of data.

Table 3. Correspondence of typical tasks and Big Data processing methods

Fraud detection	Requirements for regulations compliance	Customer segmentation	Custom offers	Customer segmentation
Machine learning	Crowdsourcing	Crowdsourcing	Statistical analysis	Visualization of analytical data
Data Mining	Statistical analysis	Data Mining	Machine learning	Statistical methods
Statistical methods	Machine learning	Statistical analysis	Predictive analytics	Simulation modeling
Simulation modeling	Data Mining	Simulation modeling	Analytical data visualization	Predictive analytics

From the analysis of the correspondence between Big Data processing and analysis methods and typical problems solved by the financial sector, it can be noted that for the same tasks, the same methods of processing and analysis of Big Data can be used. These methods are used in tools for analyzing and processing Big Data and can be sequentially or concurrently incorporated into algorithms for Big Data processing and analysis. This means that companies developing tools for working with Big Data independently decide which methods to include in the tool.

Define the criteria for selecting options for solving the issue of choosing a treatment method. Basic requirements to the criteria are:

- direct to the target;
- do not conflict with legislation;
- do not inconsistent with each other;
- compliance with modern technical standards;
- compliance with the economic level of the task.

If you set erroneous criteria for selection, then create only the form of the correct solution, which in the future can lead to errors in its use.

Criteria applicable to the selection of the Big Data processing method for solving financial sector problems are:

- regulatory documents of the Russia Central Bank, in which the recommended processing parameters for Big Data in banks, such as method, algorithm, can be prescribed. Is necessary for the banking sector, because It is impossible to work in the financial sphere, not considering the requirements of regulators, by default we believe that the criterion is fulfilled;
- time for processing and analysis. It is an important criterion for evaluating various factors that affect personal offers to customers, a quick reaction to fraudulent actions. Otherwise, the untimely offer will no longer be of interest to the client, and the fraudulent transaction will not be blocked until the circumstances become clear;
- cost of the tool containing the required method should not be greater than the estimated benefit from its implementation. This criterion is an internal limitation of the organization and comes only from the budget and preferences of the financial organization;

- load on the system when performing Big Data analysis. The negative consequences of using a particular processing method should be minimized, for example, the load on the system during processing should not be critical for the equipment, so as not to interfere with other processes of the system and not to disable the server equipment.

The last three criteria are sufficient to choose a method for Big Data processing. Let us consider them in more detail.

The formulated criteria are applicable and mandatory for the selection of a tool for Big Data analysis and processing, including certain methods of data processing and analysis, specifying the specific values specified by the internal regulations of financial organizations. Therefore, in this paper, the methods in Table 3 are ranked based on the typical solutions of the problems considered and the essences of the methods themselves.

With specific values of the criteria, they can be added to the constraints in the application by choosing the Big Data processing and analyzing method, thereby clarifying the choice of methods.

The results of the study are supported with the help of tools. On the basis of the selection criteria for the Big Data processing for solving problems of the banking sector was a program that allows you to cancel the necessary tasks and derive appropriate methods for Big Data processing.

Implemented an application that allows you to choose the most appropriate method for Big Data processing, depending on the specified type of financial sector tasks.

The application has passed approbation. The following tasks are displayed in the dialog box:

- analysis of social networks;
- the behavior of site users;
- customer credit score;
- forecasting of customers outflow;
- segmentation and customer outflow management;
- scoring;
- fraud counteraction;
- sales process optimization;
- personnel management;
- forecasting queues in offices;
- optimization of internal processes;
- management of risks;
- predictive analytics;
- marketing programs.

We will get the most suitable Big Data processing methods when selecting from the proposed list several tasks. For example, as a result of selecting tasks such as «forecasting of customer's outflow», «fraud counteraction», «forecasting queues in offices», «management of risks», Crowdsourcing will be the most suitable Big Data processing method. It's also possible to use Data Mining and analytical data visualization. The results of selecting methods for Big Data analyzing for the listed tasks are shown in Figure 1.

Fig. 1. Filled area of the method selection chart

4 Discussion

The use of Big Data is unique for each organization and depends on many factors, including the maturity of the existing information infrastructure of the organization and the availability of resources within the organization. One of the important points for the IT department head is the priority of implementing solutions in the Big Data field, identifying areas in which the Big Data technology will be of maximum benefit. As a rule, this requires a compromise between meeting management requirements, managing risks, optimizing operations and quality of customer service, which is justified in the selection criteria for both the method and the Big Data processing tool.

The use of Big Data technologies in corporate analysis, as well as the need for additional software solutions that allow for the assessment of risks in various areas of financial activity. Often when choosing the method of Big Data analysis, the cost of this method, the technical capabilities of the platform, the availability of adequate personnel are not taken into account. Availability of statistics of solved problems using analysis methods. Big Data allow to identify the most urgent and frequently occurring tasks and on the basis of this, to select an adequate solution, specialists, etc.

Conclusion

The competent use of Big Data processing methods and technologies allows financial organizations to answer the question of how to use the growing amounts of information, the speed of transmission and the variety of sources for managing changes and understanding the needs of customers, partners, and for managing risks and conducting effective work. The

criteria for selecting Big Data analyzing methods depending on the financial problems to be solved, are applicable and mandatory for the selection of the tool for Big Data processing and analysis. The main criteria are the time for request processing, the cost of the tool for Big Data processing and analyzing, the load on the system when performing Big Data analysis. Technologies and methods of Big Data processing and analysis are closely interrelated with the tools that use them. Some tools for analyzing Big Data include several methods for Big Data analyzing. In the future, the proposed approach will be developed in terms of taking into account the capabilities of the most well-known platform solutions that offer tools for Big Data processing.

References

1. W. John Foreman «Data Smart: Using Data Science to Transform Information into Insight». 53 (2014).
2. Sh. Yookpei, D. Crompton. Big data in banking for marketers. How to derive value from big data?, N, P. 328 (2017).
3. A. V. Kozlov, E. V. Rytova, S. S. Gutman, I. M. Zaychenko. The valuing of the indicator of a regional industrial development: The fuzzy logic approach 2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM). 545 – 547 (2016) DOI: 10.1109/SCM.2016.7519842.
4. A.I. Klimin, D.V. Tikhonov, A. M. Efimov. Proceedings of the 30th International Business Information Management Association Conference. Vision 2020: Sustainable Economic development, Innovation Management, and Global Growth – Madrid. 1465-1483 (2017)
5. V. Prajapati. Big Data Analytics with R and Hadoop, M., P. 238 (2013).
6. T. White. Hadoop: The Definitive Guide, 3rd Edition. O'Reilly Media, P. 688 (2012).
7. URL:<https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/>
8. T. Hastie, H. Tibshirani, J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. N, P. 126 (2016).
9. T. Pang-Ning, K. Vipin, M. Staeinbach. Introduction to Data Mining, (2015).
10. M. Hilbert. Big Data for Development: From Information- to Knowledge Societies". SSRN Scholarly Paper No. ID 2205145. P. 78 (Rochester, NY, 2016).
11. URL:[http:// www.cnews.ru/news/top/bolshie_dannye](http://www.cnews.ru/news/top/bolshie_dannye)