Sofia University "St. Kliment Ohridski"
Faculty of Economics and Business Administration

SUMMARY

Of a Dissertation on the Topic of:

AUTOMATED APPROACHES TO OPERATIONAL RISK MANAGEMENT

In fulfillment of the requirements for obtaining
The degree of "Doctor of Sciences" in
3.8 Economics

Author: **Assoc. Prof. Anton Antonov Gerunov, Ph.D.**

Sofia
2020

The dissertation on "Automated approaches to operational risk management" consists of 363 pages, of which 332 pages are main text, and it also includes a bibliography and three appendices. The main text is divided into an introductory part, five chapters and a conclusion. The study is illustrated with 141 figures and 79 tables summarizing key results. The reviewed literature is cited in the bibliographic reference, which includes 410 sources of Bulgarian and foreign authors.

Ten publications under scientific review have been made on the topic of the dissertation, eight of which are in peer-reviewed scientific journals in Bulgaria and abroad. Among the publications is a monographic textbook, which is used in the teaching of the discipline Risk Management at the Faculty of Economics at Sofia University "St. Kliment Ohridski ", a book chapter, three studies, as well as five journal articles, with one of them being Scopus-indexed. Five of the publications are in English and the remaining are in Bulgarian.

The dissertation was discussed at the Department Council of the Department of Industrial Economics and Management on 26.05.2020 (Protocol № 120/26.05.2020), following decision by the Faculty Council of the Faculty of Economics and Business Administration at Sofia University "St. Kliment Ohridski " (Protocol № 6/07.05.2020, p. 3).

Contents

Introduction

Modern economies are economies of risk. In them, the vast majority of economic activities are carried out in conditions of risk, uncertainty and uncertainty (Bullen et al., 2006). This is partly due to the information and power asymmetries that are observed and partly to the fundamental uncertainty about the future. Therefore, risk management is becoming a major task in modern economic research. Traditionally, a significant part of the research focuses on financial risks, but in recent years there has been increased interest in other types of risk, as the overall risk exposure of modern organizations is considered in its entirety. Among the main groups of risks with a potentially large effect on productivity are those that arise in the ordinary course of business. In this sense, their rational management in order to obtain maximum operational efficiency is an important task within the general research area.

In parallel with the deepening of research in the field of risk management and the implementation of key results in practice, in modern organizations there is a process of digital transformation, which is a significant change in organizational processes, structures and systems due to increased use of digital technologies and platforms (Chaffey, 2015). The trend of increasing the scope of business process automation should also cover key activities such as operational risk management. The current interdisciplinary dissertation aims to justify the possibility of automating this process by proposing a common algorithm for operational risk management, a set of methods for their quantitative assessment, as well as a technological architecture to successfully integrate them within a specialized information system.

I. Thesis Overview

Operational risk management is emerging as a leading task in the field of economics and business. As a result of the digital transformation of modern organizations, this process is also becoming a major candidate for digitization. In order for this to be successful, a common algorithm for automated operational risk management should be established, appropriate risk quantification methods should be selected and an appropriate information system architecture should be proposed to integrate all this. The main task of this dissertation is to build this system of different artifacts and results that can simultaneously support research in the field, but can be useful in scientific and applied terms. To achieve this, an interdisciplinary study was conducted in the implementation of the research tasks, which aims to expand knowledge about the approaches and methods of operational risk management in the digital environment.

1) Research Relevance and Significance

The growing complexity of the modern economy also leads to a sharp increase in the level of risk exposure of modern organizations and hence - the need for effective and efficient management of these risks (Chernobai, 2018). The increased number of business transactions and agents in globally integrated markets increases the opportunities for risk, and the digital transformation determines the transition from analog to digital risk events. At the same time, the theory and practice of decision-making in conditions of uncertainty and the management of this process are largely based on the idea of analog activities performed by human experts. This difference between the objective reality of the modern economic environment and the applicable theory for it determines the need to expand research in the field.

In addition, we take into account the growth of operational risks and the clear need to improve their management, with some authors even talking about an "explosion" of operational risks (Power, 2005). Given that this type of risk arises from four main groups of factors - people, processes, systems and external events (Chernobai et al., 2012; Leone et al., 2018), the growing importance of these factors determine the growing importance of their management. and the risks arising therefrom. Reducing the negative consequences of people, processes, information systems and the external environment and increasing the benefits of them contains significant potential for unlocking new business value and improving the competitive positioning of organizations in today's environment.

Third, technological developments make it possible to use large data sets to improve a wide range of business processes, including those for operational risk management (Davenport et al., 2012; Byrne & Corrado, 2017). This technological change also implies the use of new statistical and econometric methods for solving important economic tasks for organizations. Insufficient research on the application of methods in the field of machine self-learning and their successful integration into the overall management process in the field of risk is also an important reason for expanding research in this area.

2) Subject and Object of Study

The object of the study are the operational risks that the modern organization faces. Minasyan (2012) defines them as the risks of loss arising from inadequate or malfunctioning internal processes, people and systems, or from external events, including legal risk in this category. In this sense, the object of the study is a group of interrelated risks, each of which is a separate economic task, involving an individual approach. The subject of the research is the process of operational risk management, and of particular interest is the possibility to algorithmize this process and turn it into a set of standardized steps (activities) to be subject to automation within a specialized information system.

3) Research Objective and Tasks

The aim of the present work is to build a fully automated process for operational risk management, which can take advantage of a wide range of quantitative assessment methods and be supported by a relevant specialized information system. Automated risk management has the potential not only to improve organizational efficiency by improving the quality of decisions and the speed of their implementation, but also to reduce costs by facilitating some of the work of risk managers and experts who support them. To achieve this goal we perform the following research tasks:

1. Review and critical analysis of risk management approaches and methods with a focus on operational risk management;

2. Derivation of the main economic, social and technological trends that should be taken into account in the automation of the management process;
3. Operationalization of the definition of the concept of operational risk in a way that can be applied within an information system;
4. Derivation of a general algorithm for operational risk management, which is subject to automation;
5. Analysis of the methods for assessment of the operational risk in tasks of classification type and derivation of the most optimal from the point of view of the forecast accuracy algorithms, which can serve the needs of the algorithm;
6. Analysis of the methods for assessment of the operational risk in tasks of regression type and derivation of the most optimal from the point of view of the forecast accuracy algorithms, which can serve the needs of the algorithm;
7. Construction of a reference architecture of a management information system, which will allow the application of the proposed algorithm and methods in an automated way, as well as to ensure seamless inclusion in the business processes and the general organizational architecture;
8. Analysis of the social and ethical aspects of the implementation of an information system for operational risk management with the possibility for autonomous decision-making.

The fulfillment of the set research tasks leads to the fulfillment of the general goal of the dissertation, determining the accumulation of sufficient results to allow the construction and implementation of an automated process for operational risk management.

4) Research Hypotheses

The main thesis of the presented study is that the process of operational risk management is subject to full automation by applying an algorithmic management approach, a set of algorithms suitable for analysis of large data sets and building an appropriate information system, building on reference architectures known in the literature. (ex. Klein et al., 2016).

The working hypotheses of the study are:

1. All activities performed by human experts in the operational risk management process may be automated.
2. The application of classification algorithms in the field of machine self-learning for quantitative risk assessment can improve the forecast accuracy compared to the traditionally used econometric methods and hence lead to higher quality of results and economic value.
3. The application of regression algorithms in the field of machine self-learning for quantitative risk assessment can improve the forecast accuracy compared to the traditionally used econometric methods and hence lead to higher quality of results and economic value.
4. The use of several criteria for assessing the level of risk within an automated algorithm will lead to grouping (clustering) and not to divergence of results. In this way the reliability of the proposed system is guaranteed.

The testing of the working hypotheses is performed according to their specifics. Hypothesis 1 is tested qualitatively, applying the approach of design science (Hevner et al., 2004; Arnott & Pervan, 2016). Creating an artifact that meets certain requirements rejects the null hypothesis in this case. Hypotheses 2 and 3 are tested quantitatively by comparing the forecast accuracy of the methods proposed here with the forecast accuracy of the naive prediction and that of the classical regression econometric methods. Hypothesis 4 is tested with a combination of quantitative and qualitative methods, taking into account the correspondence between the different criteria for risk levels and taking into account the statistical definition of anomalous (extreme) value.

5) Scope of Study

The study covers twelve main situations of operational risk to which the proposed general algorithm is applied, as well as a set of specialized methods in the field of machine self-learning. These twelve situations are essentially twelve different tasks, and they are described in the corresponding data set. They relate to operational risk management in the following situations:

1. Conducting a direct marketing campaign
2. Conducting credit card operations
3. Granting loans
4. Relationship management with external partners
5. E-commerce activities
6. Excessive absences from work
7. Online communication
8. Valuation of asset prices (real estate)
9. Sharp changes in market demand
10. Support ticket processing
11. Marketing communication through social networks
12. Demand management in an e-shop

From a time point of view, the data are predominant from the last 10 years, as is true for 10 of the 12 arrays examined. The application of the proposed approaches and algorithms to a relatively large set of operational situations in both the analog and digital worlds, provides their flexibility and the ability to transfer to a wide range of other tasks in the field of operational risk.

6) Methodology

This dissertation uses both general scientific and specialized methods to perform the research tasks. The general methodology is based on a systematic approach and follows general scientific principles such as objectivity, transition from concrete to abstract, concretization and unity between theory and practice. Within the literature review and critical assessment of the existing typologies of risk, as well as the methods and approaches for its management are the used methods of analysis and synthesis, as well as inductive and deductive methods for reaching the main conclusions. This methodological toolkit is suitable for this type of tasks, as it allows an overview and summary of a wide range of literature sources, and analysis and synthesis allow the derivation of new scientific results from the analysis of existing research. The systematic and interdisciplinary approach, on the other hand, contributes to the unification and integrated understanding of results coming from different disciplines and applied fields - economics, risk management, econometric and statistical methods and management information systems.

When building the general algorithm for operational risk management, as well as when deriving the reference architecture of an appropriate information system, it is appropriate to use not only analysis and synthesis, but also skills for designing new artifacts. For this reason, the approaches and methods of design science are used here (Hevner et al., 2004; Arnott & Pervan, 2016), with special attention paid to prototyping methods. This approach allows to build new useful artifacts with high scientific and applied value based on a set of certain requirements. These requirements can be either directly following the function of the artifact (eg requirements for the management information system to perform any activity of a control algorithm) or non-functional requirements following from other sources (eg technological constraints or ethical considerations). In this sense, the proposed algorithm and reference architecture of the system are the result of a combination of general scientific and specialized methodology of design science.

The analysis of the different types of statistical algorithms for machine self-learning and the selection of the most accurate among them for the needs of automated control presupposes a specialized methodology. For this purpose, standard methods and approaches from the field of statistics and econometrics are used. Both well-known and relatively standard econometric models (correlations, linear and logistic regressions, etc.) and a set of advanced methods in the field of machine self-learning (neural networks, random forests, Bayesian methods, machines) are used in the modeling of operational risk situations. with supporting vectors, etc.). To assess the importance and contribution of certain variables to the models, statistical hypotheses are derived and they are formally tested by calculating test statistics and exact levels of significance. Statistical criteria for forecast accuracy were defined and applied for the evaluation of the algorithms and they were compared quantitatively. Quantitative statistical methods allow precise testing of the hypotheses (mainly Hypotheses 2 and 3, and partly - Hypothesis 4), while expanding the methodological knowledge of the optimal methods for operational risk management. All calculations are done via the R language for statistical computing and its associated packages (Kuhn, 2008).

7) Data Sources

The information provision of this study is key to its successful implementation. The quantitative data used in the dissertation are of three main types. The former are statistics coming either from a primary source or cited by other authors. In both cases, their origin is clearly marked. The latter are the data used to assess operational risk situations, which is generally an analysis of secondary data. The third is simulated data, using the built-in random number generator in the programming language R. The risk management process has a number of characteristics that pose challenges in its modeling. First, it is highly information-intensive and requires a certain scope and level of data quality. Moreover, the successful application of some of the potential quantification methods presupposes the existence of relatively large samples that will allow the correct and unbiased calculation of their parameters. Third, operational risk has a number of different manifestations in different situations, which can be described by data of different format and scope, and their format, structure, type and type cannot be standardized - ie. there is not one task for evaluation and management, but a series of different tasks. All this implies a more precise selection of the data used.

The majority of the analyzes performed use data for various online and offline operational risk situations provided by researchers in this and similar fields. The full set of data used is presented in Table 1. Apart from quality assurance and the required sample size, these data are most often tested in the analysis and development of quantitative methods. This allows the results of the dissertation to directly build on established developments, but also to be compared with the results obtained by the respective authors, assessing whether the algorithms proposed here represent an improvement over the originally used ones. This is the case in each case.

Table 1: Data sources

№	Operational Risk Situation	Data Source
1	Conducting a direct marketing campaign	Moro et al., 2014
2	Conducting credit card operations	Yeh & Lien, 2009
3	Granting loans	Eggermont et al., 2004; Hofmann, 1994
4	Default probability of external partners	Zieba et al., 2016
5	E-commerce activities	Sakar et al., 2018
6	Excessive absences from work	Martinian et al., 2012
7	Online communication	Fernandes et al., 2015
8	Valuation of asset prices (real estate)	Yeh & Hsu, 2018
9	Sharp changes in market demand	Ferreira et al., 2016
10	Support ticket processing	Amaral et al., 2018
11	Marketing communication through social networks	Dehouche & Wongkitrungrueng, 2018
12	Demand management in an e-shop	Chen et al., 2012

8) Utility and Novelty of Results

The usefulness of the research derives from the results achieved in scientific and applied science, and it should be noted that they can be used in more than one field of scientific interest. Their applicability is in the fields of economics and the science of risk management, business management and the implementation of management information systems. First, the developed general algorithm for automated management allows to apply a unified management approach to all situations of operational risk, in order to standardize and optimize this business process in the modern organization. Moreover, the algorithm is derived from the explicit idea that it can be applied as part of the efforts to implement tools of digital transformation and therefore provides opportunities for partial or complete automation of both core activities and the process as a whole. Therefore, this approach can be used both as a framework or basis for future research in the field of risk management and digital transformation, and in a scientific-applied context in the implementation of such solutions in the structures of the private and public sector.

Second, the analysis of the algorithms for quantitative risk assessment shows which methods are appropriate to apply in solving problems in the field of risk management. Improving the forecasting accuracy of various analytical activities and systems has the potential to improve organizational decision-making and hence generate performance improvements, cost reductions and potential new competitive advantages. In addition, the results of the considered algorithms can be relatively easily generalized and transferred to other fields relying on the analysis of large data sets, which is useful for the relevant research and applied developments in these fields. From a methodological point of view, the comparison of different algorithms and methods for solving similar problems directs researchers to the optimal ones, which allows both their wider use and potential development.

Third, the derived reference architecture of a management information system for operational risk management in scientific terms can be used as a starting point for future research, and in application - as an artifact to facilitate the process of developing and implementing such a solution. Fourth, within the dissertation a number of definitions, parameters and criteria are derived and operationalized, the explanation of which has a pronounced scientific and applied benefit. Appropriate criteria for automatic selection of an optimal algorithm between different alternatives are outlined. The application of algorithms for unsupervised self-learning for solving tasks in the field of risk is also shown, and on the basis of a consensus criterion their combination is demonstrated. In the literature, such approaches are limited mainly to risks of financial fraud and information risks, but the dissertation emphasizes how they can be applied to operational risks arising from consumer behavior.

9) Study Limitations

The presented study also has some limitations that should be taken into account so that the results obtained can be correctly interpreted within this framework. In the first place, it should be borne in mind that a large but still limited set of situations involving operational risk is being examined. The dissertation does not comprehensively cover all possible operational risks, nor all possible corresponding data sets that describe them. Therefore, it is possible in the application of the algorithm, the set of methods for quantification and in the reference architecture of the systems to a new and unexplored situation to require their adaptation or extension. Secondly, in order to ensure comparability between the different algorithms, the same values of similar parameters were used in their training. This is especially true when evaluating DBSCAN and LOF unsupervised self-learning algorithms. Although this provides an opportunity to compare the given algorithms, it is appropriate to carry out a formal calibration process that iteratively seeks the optimal values of these parameters for a given criterion (eg percentage of potentially risky observations). It is possible that when changing the parameters of some models, small changes in the obtained results will occur, and it is appropriate to study in more depth.

Third, the relatively limited scope of the business rule management module should be taken into account. He is committed to the automation of actions performed to manage operational risks and in its current form assumes the existence of specific rules (business logic) for performing certain activities in identifying potential risk (eg when detecting a risky transaction in real time to send a request to the processing system for its termination). This implies an individual definition of the rules within each organization and for each type of operational risk. Although this is the standard approach at the current level of technological development and organizational maturity, with the development of technology this module can be expected to be replaced by a more complete version of limited artificial intelligence, so that at least some of the rules are generated automatically. This question is also particularly important in terms of how the system would react in the event of severe external shocks. These shocks can be conditionally divided into two groups. In the first, we observe accidental shock or extreme deviation, and this type of shock can be ideally identified by the proposed solution. The second type of shocks are essentially the first observations after switching the modeled system to a new mode and suggest some time to adapt and train the models to this new baseline level. The existence of in-depth business rules or the application of specialized artificial intelligence can shorten this adaptation period.

Fourth, a relatively limited assessment of the ethical aspects of the introduction of the automated management information system can be noted as a potential limitation. The assessment of ethical aspects in the case of autonomous digital agents is still in its infancy and although it is in line with the strategic guidelines for the development of artificial intelligence within the European Union (European Commission, 2019), methodologies and approaches are relatively limited. and are subject to further development and expansion.

10) Directions for Future Work

The guidelines for future work in the field of the researched issues largely follow from the identified limitations. First of all, it is appropriate to expand the scope of the considered operational risk situations, maintaining the balance between situations typical of the analog world (physical or present business processes) and those typical of the digital world (digital business processes). Considering more situations and applying the derived algorithm and quantitative evaluation methods to it has the potential to lead to further validation of the results presented here or to upgrade and expand them so that they are even more useful for the scientific community and business. .

Secondly, there is significant potential to work towards expanding the business rules management module. This is precisely the activity that is most difficult to automate due to the specifics of each individual situation of operational risk and the way in which different organizations address these situations. Moreover, the full set of business rules is not always clear, and some of them may be in implicit form. Therefore, it is realistic as a first step in the work to expand this module to analyze data on actions taken against a given operational risk and on this basis to start automatic derivation of rules (Raad et al., 2017; Mutlu et al., 2018). This approach has the potential to automate the generation of rules, and at a later stage a more complete agent with limited artificial intelligence can be integrated. A third direction for future work is the development of a methodology for economic and ethical evaluation of management information systems with the ability to make autonomous decisions. Given the growing importance of information assets for the competitive positioning of modern organizations and the ever-increasing scope of information systems, it is appropriate to derive a formal methodology for assessing their economic, social, technological and ethical aspects.

II. Structure of Dissertation

The dissertation is structured in five chapters, introduction, conclusion, bibliography and three appendices. Each of the chapters focuses on achieving one or more of the set objectives, presenting summaries and new results in response to the need for better operational risk management.

Table of Contents

The first chapter is a review of the literature in the field of risk management, with a subsequent focus on operational risks. The main typologies of risks in general and operational risks in particular are considered, and a critical assessment of each of them is made. The quantitative and qualitative methods for operational risk management are reviewed, as well as the general strategies and good practices for this process. This chapter argues that although quantitative methods offer the highest precision and efficiency of the process, in their current version they are overly dependent on the availability of a relatively large number of qualified human experts to be successfully applied. This emphasizes the need for full or partial automation of this management process with the help of a specialized information system.

The second chapter focuses on a review of new developments in the field of operational risk management and, based on the literature review, it outlines four main trends in current research. First, there is an increased focus on industry-specific and specific operational risk management situations, which partly reflects the global mega-trend to customize goods, services and processes. Secondly, there is a deepening of research in the literature regarding the risk and penetration of information and communication technologies, as the number of studies in the field of information security risks is growing rapidly. The current research also shows the introduction of new and advanced methods for assessing operational risk and making adequate management decisions in conditions of uncertainty. Lastly, there is a tendency to achieve an integrated view of the overall risk exposure of the organization and a reduced focus on the synthetic separation of certain risk groups.

In the context of new developments in the field of the global environment, technologies and research in the field of operational risks and based on established methodologies for analysis, this chapter also presents the general algorithm for automated management of operational risks. It consists of eight steps, separated into five main management stages - problem definition, information support, training of operational risk management model, application of the model and management actions. The stages presented in this way take into account the main phases in the management process, but also allow algorithmization of actions so that they can be applied within a specialized information system.

Chapter Three aims to identify appropriate algorithms for quantifying operational risks in cases of discrete choice. For this purpose, 136 different algorithms from the field of statistics and machine self-learning are tested, and they are applied to solve five classification problems in the field of operational risk. These situations and their corresponding databases report the following cases involving risk management: conducting a direct marketing campaign, conducting credit card transactions, granting loans, managing relations with external partners, e-commerce activities. data sets, these 136 algorithms were calculated and based on their predictive accuracy, measured by the area under the performance curve, the ten methods with the highest accuracy were selected.

Among the various tasks, a predominantly good presentation of the methods from the random forest family is observed, and from the classical econometric methods high accuracy is achieved by the linear discriminant analysis. Additionally, a complexity measure is reported here, taking into account the computational resources needed to evaluate each of the algorithms. The results show that the algorithms with the highest forecast accuracy are not necessarily characterized by the highest resource consumption. This gives grounds to argue that the choice of the optimal algorithm should be made by achieving a balance between its costs, measured by the necessary computational resources and its benefits, measured by its forecast accuracy.

Chapter Four discusses and identifies appropriate algorithms for quantitative risk assessment in cases where the target task variable is continuous. This presupposes the testing of regression algorithms, as here again both traditional econometric methods and those in the field of machine self-learning are analyzed. A total of 109 different algorithms are tested, and they are applied to solve five situations from the field of operational risk with regression character. These situations include the management of excessive absences from work, online communication, valuation of assets (real estate), abrupt changes in market demand and signal processing by customers.

Again, all algorithms are evaluated for each of the data sets and a measure of complexity is reported, taking into account the resource intensity of each of the considered methods. Here, too, there is a very good representation of the methods of the random forest family, but also of neural networks. In contrast to the classification problems, in this case the classical econometric methods of linear regression have a significantly lower performance compared to methods in the field of machine self-learning. In regression tasks, we again observe a relatively weak relationship between forecast accuracy and resource intensity, which allows to make an optimization choice between the benefits of a given algorithm and the cost of its evaluation.

Chapter Five discusses the possibilities for integrating the proposed common algorithm for operational risk management and the corresponding methods for quantitative risk assessment within a specialized information system. To begin with, the types of information systems and different reference architectures of the group of management information systems are considered. Based on established architectures and good practices, with the help of methods from the design science a reference architecture of an information system for automated operational risk management is derived. The system contains four main subsystems that aim to automate the activities proposed in the algorithm - these are the storage and processing subsystem, the modeling subsystem, the analysis subsystem and the management subsystem. An additional three horizontal modules (security and access control, system management and integration) complement the system and ensure the implementation of non-functional requirements to it.

There is also a dedicated module for automation of the process of determining the target variable using four algorithms for unsupervised self-learning. A consensus anomaly criterion is derived based on these four algorithms, with observations classified as anomalous by three of them defined as risky and those classified by four as high-risk. This approach has been applied and tested in two additional new situations of operational risk - in marketing communication through social networks and in demand management in an e-shop, and its usefulness is shown. The expansion of the information system with possibilities for autonomous decision-making brings it closer to some characteristics of the limited artificial intelligence. Therefore, compliance with ethical requirements for autonomous agents was assessed, and the main conclusion is that the proposed system has a very good level of compliance with these requirements.

III. Main Results of the Dissertation

Chapter 1: Economic Risks and Approaches to Their Management

The last two centuries of economic history have been characterized by the exponential growth of the world economy - both in absolute terms and per capita in a steadily growing population. Although these factors are often perceived as unequivocally positive, the growing interconnectedness and dependence of different sectors and the accelerated pace of innovation in economic activities determine the evolution of the modern economy as a complex and difficult to predict system.

Although economic risk has always been an integral part of economic activities, its effects are much more pronounced in the context of complex systems with accelerated information processing such as many sectors of the modern economy. Therefore, it is not surprising that the formal study of risk has become increasingly important in the last century and currently risk management is not only the subject of academic research, but also an important part of modern management practices in the private and public sectors. The scientific understanding of risk is strongly related to the field of research and application, but in any case is largely determined by the modeling and management of random events through the tools of probability theory.

While the definition of risk is relatively stable and widely accepted, the definitions of the different types of risk vary considerably. Focusing on operational risk, we should note that it is defined differently by different authors. Some of the alternative definitions are presented in Table 2: Definitions of operational risk.

Table 2: Definitions of Operational Risk

Year	Author	Definition of Operational Risk
1993	Group of Thirty	Uncertainty related to losses resulting from inadequate systems or controls, human error or management problems.
1998	Crouhy et al.	The risk that external events or deficiencies in internal controls or information systems will result in a loss - whether the loss is expected to some extent or completely unexpected.
1999	The Commonwealth Bank of Australia	All risks with the exception of credit and market risk, which would lead to fluctuations in revenues, expenses and the value of banking activities.
2000	Jorion	The risk arising from human and technological errors or accidents.
2000	King	A measure of the relationship between the business activities (processes) of the organization and its results.
2001	Crouhy et al.	The risks associated with the activities of a business.
2001	Basel Committee	Risks arising from inappropriate or failed internal processes, employees or systems, as well as from external events.
2002	Lopez	Any non-quantifiable risk faced by the bank.

2003	Securities and Exchange Commission	Risks of loss resulting from failed controls within the organization, inexhaustibly including unidentified overdrafts, unauthorized trading, fraud in trading or support functions, inexperienced personnel, and unstable or easy-to-access computer systems.
2005	Deutsche Bank	The potential for realization of losses arising from the actions of employees, contractual relations and documentation, available technological solutions, non-functioning of the infrastructure, incidents, external influences and customer relations.
2005	Vinella & Jin	The risk that an activity will not achieve one or more of its operational objectives, and the activity may be related to people, technology, processes, information or infrastructure that supports business processes.
2006	Basel Committee on Banking Supervision	Risks arising from inappropriate or failed internal processes, employees or systems, as well as from external events.
2009	Solvency II (Directive 2009/138/EC)	The risk of a change in value caused by the fact that the realized losses from failed internal processes, people and systems, or from external events (including legal risks) differ from the expected losses.
2012	Минасян	The risk of loss arising from inadequate or malfunctioning internal processes, people and systems, or from external events, including legal risk.
2012	Chernobai et al.	Unsystematic risk associated with four main groups of factors: (1) People; (2) Processes; (3) Systems; (4) External events.
2018	Leone et al.	The risk of loss resulting from inadequate internal processes, human error or failure of certain systems.

Chernobai et al. (2012) emphasize that operational risk is unsystematic and strictly organization-specific. In this sense, its diversification is very difficult, if not impossible. Leone et al. (2018) provide a detailed review of alternative definitions of the concept of operational risk, based on banking practice, where operational risk is defined as the risk of loss resulting from inadequate internal processes, human error or failure of certain systems. Jorion's (2000) definition is similar. To this definition, the Bank for International Settlements adds the risk arising from external events (see BIS, 2001). In practice, a much broader understanding of operational risk is required, with Deutsche Bank (2005) defining it as the potential for losses resulting from employee actions, contractual relationships and documentation, available technological solutions, infrastructure failures, incidents, external influences and customer relationships.

Vose, 2008 presents a detailed methodology for quantitative risk management, and it is worth noting that it explicitly outlines the two key roles of those involved in the process:

- **Analyst** - an expert who models and assesses the risks by building a formal model, producing results in the service of formal management;
- **Process manager** - most often a manager or group of managers responsible for managing the risk exposure.

Chronologically, the management process begins as the process manager formulates organizational problems and context and lists the potential risks and opportunities for their management. This can be done both by a specific individual and by groups of managers and experts depending on the needs and practices of the organization. As an important point here we consider the need to choose quantitative criteria for selecting appropriate models, approaches and strategies for risk management. Everything so far can be considered as input data for the subsequent analytical process. The analytical process consists of the following main steps (Vose, 2008, p. 5), which are usually performed by an expert or a team of risk management experts (see ibid., Pp. 23-26, as well as Kroenke et al. , 2012; Larose et al., 2014; Lu, 2018):

- **Review and analysis of available data -** the first step is review and analysis of available data. This most often involves descriptive (descriptive) analysis, as well as deriving the relationships between the various variables that are relevant to building the overall risk model. The relationships can be traced both visually (eg by heat map) or analytically (eg by correlation matrix).

- **Model design -** as a next step, experts should build a common architecture and structure of the model that will be used for risk management. Here it is appropriate to choose both the type of model (level of formality, use of empirical data, simulation) and the specific algorithms to be used (eg correlation or regression analysis, machine learning methods, optimization algorithms, etc.) .

- **Model building -** this step involves building the model, as quantitative models are often built and subsequently executed in a specific software environment. In this step, the approaches selected in the previous one are performed using the various appropriate algorithms.

- **Derivation of probability distributions of parameters -** after the creation of the model the parameters included in it are evaluated. When using the parameter estimation approach based on previous data, these parameters are calculated and their level of uncertainty is determined (most often in the form of a confidence interval). Apart from the evaluation of the main characteristics of the variables, it is sometimes necessary to derive their statistical distribution, summarized as a function of distribution or probability density.

- **Conducting simulations -** using the calculated estimates and the derived probability distributions of the key variables, the analyst can make simulations (eg by the Monte Carlo method) to take into account the expected distributions of results, as well as to estimate the required percentages of expected risk realizations for the needs of its management. At this stage, a scenario analysis can be performed to show the sensitivity of the results to changes in some or all of the parameters included in it. Such an approach allows to quantify the model uncertainty that

- **Validation of the model -** after the construction of the quantitative model, it should be validated, and this validation can be compared with other models (eg on the basis of information criteria, accuracy, explanatory power, etc.) or on the basis of expert evaluation of the obtained results. The results of the validated model can be used in subsequent steps of the management process.

- **View results -** in this step the results obtained from the model and their benefit and application within the general process are discussed. Both analysts and process managers participate in this review and from an organizational point of view this is a moment not only of approval of the obtained results, but also of communication and smoothing of potential differences.

- **Report on the results of the model -** after reaching a common understanding of the results obtained from the modeling, they are formed and presented in an appropriate format in a report. The aim is for the visualizations and analytical presentation to support the management process to the maximum extent.

- **Model support -** the last important step of the analytical process is the maintenance of the model, which includes both updating of the data and estimates of the used parameters, as well as its technical support. The latter is especially true when using large models that require a complex computing infrastructure.

As an alternative approach to operational risk management, the use of quantitative data and their modeling using statistical and econometric methods is required. A number of authors (Vose, 2008) recommend this approach as more reliable and formally traceable. We note that the designation quantitative approaches in practice contains a wide range of different analytical methods using a variety of data sources. Quantitative management of operational risk is largely determined by the requirements for its management in the financial sector, but it is worth noting alternative approaches in other areas of the risk economy (Galloppo & Regora 2011; McNeil et al., 2005; Frachot et al. , 2001). To this end, we first take into account the requirements of the Basel Accords and then outline the principles of using Monte Carlo methods in the assessment of operational risk (Akkizidis & Kalyvas, 2018).

In research, the standard economic approach is to emphasize the principle of methodological individualism and use the basic results of utility theory to analyze risk decision-making problems (see, e.g., Damodaran, 2007). Much of the literature is dominated by research and analysis of surrounding financial risks, which are often placed within the regulatory requirements for financial institutions (eg Eliot, 2012; Leone et al, 2018; Guegan & Hassani, 2018). This focus leads to the fact that financial risk management is largely refined, with a focus on its application and evolutionary development, driven primarily by regulatory requirements. It is noteworthy that our scientific literature also reflects the increased attention on financial risk management, which leads to a number of scientific and scientific-applied contributions of native scientists.

There is a marked tendency in the scientific literature for a synthetic distinction between the different types of risks and the respective specific strategies for their management. This gives rise to various classifications of risk groups, some of which are presented in the present study (Frame, 2003; Crouhy et al., 2006; Dionne, 2013; Balabanov, 1996; Zafirova, 2016, etc.). We recognize that such an approach is useful from an analytical point of view, but from the point of view of research and applied research it is often problematic. In particular, we emphasize the high correlation between different types of risks and the significant overlap between some of them. This presupposes an integrated consideration of the risk exposure of the modern organization and a unified approach to its management. The main growing risk is operational risk, which has significant potential to lead to critical losses, while some aspects of its management are subject to significant improvement. In this sense, operational risk management is emerging as an important and topical topic of research.

Here we have considered a standard definition of operational risk as arising from people, processes, information systems and external events (Chernobai et al., 2012), as this definition is imposed by consensus and serves as a basis for various taxonomies of operational risks (Embrechts et al. , 2003; Jarrow, 2008). In the context of the increased complexity of the economic environment, the types of risks faced by organizations and individuals, as well as their intensity, are also changing. As an example, Chernobai et al. (2018) note that with the growing complexity of the financial environment, operational risks for banking financial institutions also increase significantly. This trend further enhances the relevance of research in this area.

In practice, the use of quality methods for managing operational risks is largely necessary because they are easy to apply and can be adapted to a wide range of different situations (Pritchard, 2014). These advantages are offset by their significant inaccuracy. Most remarkably, the qualitative labels for the size and probability of the risk are not semantically equivalent and in this sense are incomparable (eg two risks with a "medium" probability do not have the same probability of realization). They also do not allow a formal optimization decision to be made comparing the benefits and costs of their management, nor, in general, to examine the uncertainty in these assessments.

Therefore, a number of researchers (see eg McNeil et al., 2005; Crouhy et al., 2006; Vose, 2008; Akkizidis & Kalyvas, 2018) prefer the quantitative management of operational risks, which includes specific numerical values of key parameters and derivation. of formal indicators for the level of risk - eg beta coefficients, standard deviations, Sharpe coefficients, metrics for value at risk or expected extremely severe loss. This approach is characterized by the fact that it is particularly intensive in terms of its information needs - it requires relatively long time series of data or relatively accurate expert assessments of the basic parameters.

Although quantitative approaches are considered to be the most advanced risk management methods at this stage, they suffer from a number of shortcomings. In particular, they cannot always identify specific sources of risk, there is no uniform methodology for their application and they do not lead to specific recommendations for risk management strategy or activities. Moreover, at this stage, quantitative risk assessment is highly dependent on the use of highly qualified experts who, in collaboration with management staff, assess exposure and develop strategies to address uncertainty. Given the growing importance of information systems and the growing need for faster (even real-time) decisions, this approach is unsatisfactory. In this sense, there is a clear need to improve and automate approaches to managing operational risks so that they meet the needs of modern organizations.

Chapter 2: New Approaches to Managing Operational Risk

The rapid development of technology over the past two decades and the corresponding changes in the socio-economic context of modern organizations have significantly changed the possibilities for, but also the need for, new models for operational risk management. The review of the main trends in the field so far clearly outlines the direction of development of new approaches, noting also the lack of a full-fledged and universally accepted approach for digital (digital) transformation of organizational risk management. On the one hand, we see a number of unlocking technologies that should be applied to this process. First of all, we note the presence of large data sets ("big data"), which are generated constantly and in real time by a number of operating and transactional systems. The development of technological approaches for their collection, storage, processing and analysis provide a number of opportunities for businesses to offer new consumer value and even co-create value, but also reveal new risks and challenges on an unprecedented scale. The exponential growth of data and the accelerated introduction of digital products, services and processes contribute significantly not only to the digital revolution in modern organizations, but also to the expansion of their risk exposures.

Technological developments are accompanied by socio-economic ones. Most notably, we note the growing division of the labor market, with a shortage of highly qualified staff supporting the process of transferring low-skilled work to automated and robotic systems. On the other hand, this process calls into question the jobs of the low-skilled workforce. Additionally, we are seeing an increase in the complexity of business processes and the complexity of the management of modern organizations under the influence not only of IT breakthroughs, but also of globalization and the use of global supply chains and value creation. These main engines significantly change the context of operational risk management, while giving the opportunity, but also set the imperative for the automation of this process.

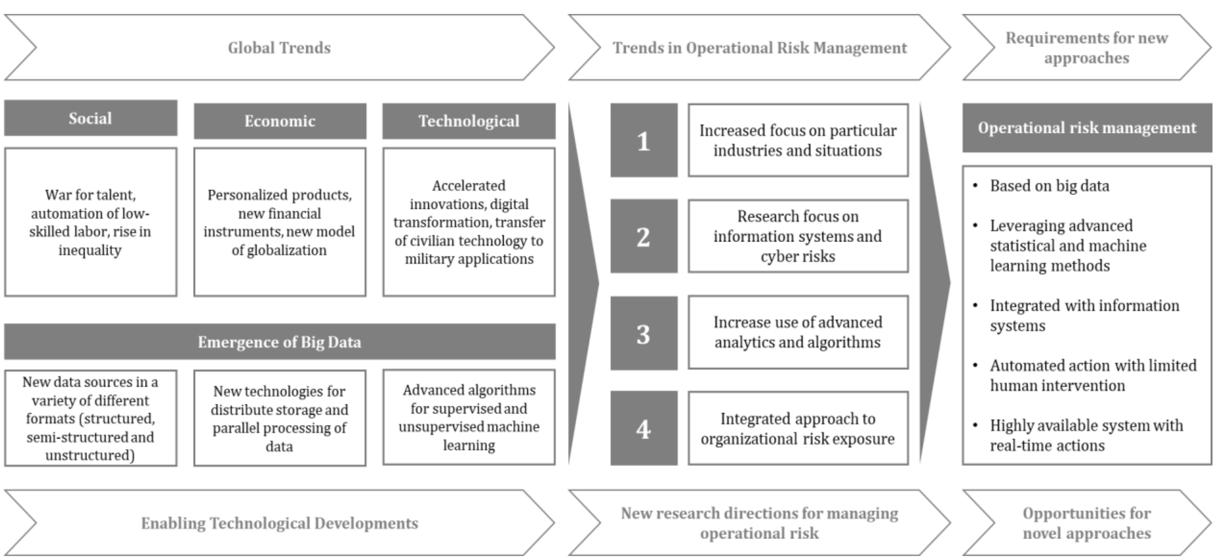

Figure 1: New Developments and Requirements for Operational Risk Management

This trend is becoming increasingly visible in the new research areas and approaches in the field of operational risks. Based on the presented review, we report four main trends in the study of operational risks. First of all, there is an increased focus on specific application fields and problems, which makes the risk management process personalized for the given problem in a way similar to the personalization of digital products and services to the individual user. Secondly, we note that information security risks are becoming a major class of operational risks, due to the accelerated penetration of information systems in almost all aspects of modern business.

Third, we consider the accelerated research, testing and partly the introduction of advanced algorithms for data analysis and risk assessment, as these algorithms are adapted to function effectively within the modern IT architecture, allowing speed, scalability and distributed computing. Fourth, the tendency for an integrated view of the overall risk exposure of the modern organization stands out. Unlike the previous focus on the precise analytical division between the types of risk groups, modern approaches rely on a general holistic overview of risk, taking into account the high correlation between different risk groups in a real environment and emphasize that their division is often difficult. The trends described in this way set new opportunities and requirements for modern approaches to operational risk management and outline the contours of a new paradigm in the field. Figure 1: Main trends and requirements for new approaches to operational risk management shows schematically how the change in context, technological conditionality and new research lead to new requirements for the modern management process.

We note five main requirements for new approaches to operational risk management, as follows:

- **Based on large data sets** - the exponential growth of available information suggests that organizations have larger data sets that contain knowledge of their operations. The expansion of the information set is possible and desirable in order to more precisely define and assess operational risks. The added value of analyzing large data sets is already an established fact and operational risk management should make the most of available organizational knowledge. Moreover, there is a technological opportunity for a wide range of new data in various formats to be successfully integrated into the management process so as to generate business value (McAfee et al., 2012). The cost of such integration is also significantly reduced (Byrne & Corrado, 2017), which makes it financially viable for many organizations.

- **Use of advanced statistical algorithms** - we noted the availability and introduction of advanced methods in the field of machine learning for the needs of business analysis, decision making and risk management. A number of studies (eg Makridarkis et al., 2019) show significant differences between the results of different statistical algorithms, arguing that the most optimal among them are significantly better than their runners-up. Since even a small improvement in the predictive accuracy of operational risk management algorithms can lead to a significant increase in the value generated by them, it is necessary to fully explore the optimal algorithms for different types of sub-tasks of operational risk management and they to be applied in practice.

- **Integrated with information systems -** a key change in the operations of modern business is rooted in the growing use of automated information systems and the need to integrate them in order to obtain a unified management environment for the organization. It is the decision-making algorithms in the context of risk and uncertainty that would benefit most from the integration with the information arrays of a given structure, from where the management algorithms can be directly fed with data. Additional integration with operating and transactional systems will allow automated action to be taken as a result of risk analysis, which can speed up this process and deliver significant value (see Raghupathi & Raghupathi, 2014).

- **Automatic actions with minimal human intervention -** digital transformation involves the passage of an increasing number of processes and solutions beyond human control and their automation with the help of appropriate systems. In fact, automated decision-making is emerging as a leading practice that saves time, generates consistent decisions, and helps overcome an individual's cognitive limitations. Their growing application (Battaglini & Rasmussen, 2019) shows that automation has significant potential to support the activities of modern organizations and lead to better results.

- **Ongoing and real-time process** - the availability of an automatic data flow to a risk management and decision-making algorithm means that the available information available to this module is updated in real time. This also requires real-time monitoring, analysis and assessment to ensure that the most up-to-date data is included in the risk management decision of each specific action or transaction (Davernport et al., 2012). This is especially true for a wide range of information security risks (e.g. cyber attack), in which seconds of delay can predetermine the success or failure of preventive and corrective actions.

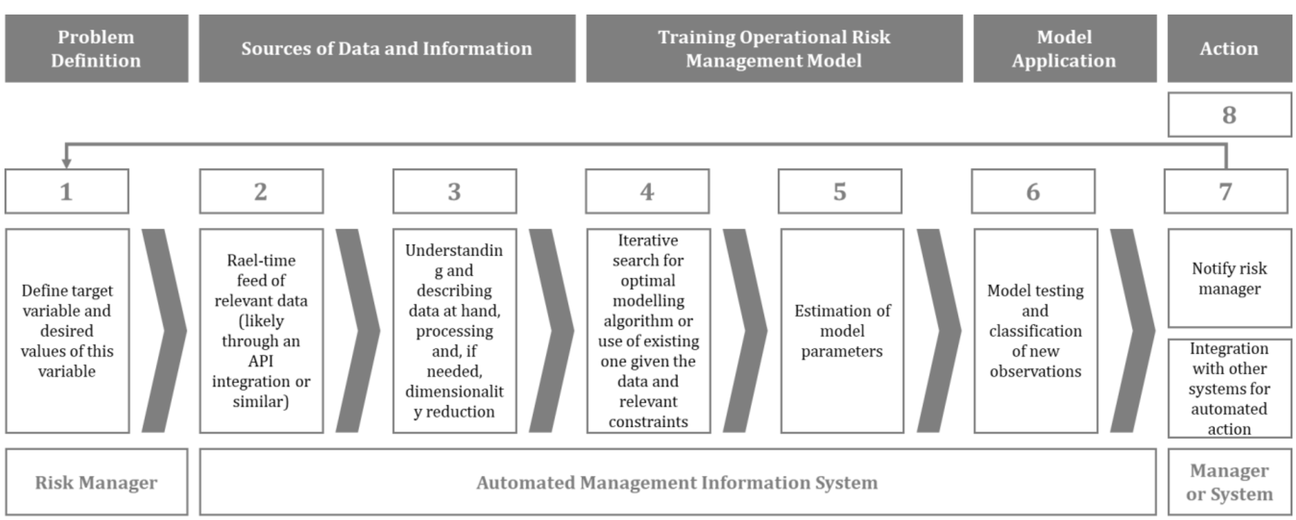

Figure 2: General Algorithm for Automated Operational Risk Management

In this sense, the operational risk management process can no longer be periodic compared to a pre-defined schedule in the organization's policies, but should be constant and in real time. Real-time decision-making is also emerging as an active and promising area of research and applied risk research in a wide range of different fields (see, e.g., Sanders et al., 2018). The derived requirements for the new approaches to operational risk management can be used to extend the classical approaches to its management (eg Vose, 2008), for this purpose automating the steps that require human intervention and integrating a formal methodology for derivation. of knowledge from data such as CRISP-DM (Wirth & Hipp, 2000; Azevedo & Santos, 2008; Bosnjak et al., 2009; Schaefer et al., 2018; Huber et al., 2019). By emphasizing analytical components and updating the approach to operational risk management through a synthesis of scientific results from the last decade, we arrive at a new approach to operational risk management (see Figure 2). This innovative approach builds on current scientific advances and best practices and adapts them for use within automated systems for analysis and autonomous decision-making.

The general algorithm for automated management of operational risks consists of 8 main steps (processes), which include the full management cycle of operational risks - from problem definition and data collection, through analysis and measures, to their application. The steps are divided into five main groups of activities that largely follow the usual management cycle (Haimes, 2015; Vose, 2008; Crouhy, 2006; Sadgrove, 2016; Hopkin, 2018; Zafirova, 2016; Tsanevska, 2017):

- Definition of the problem
- Information support
- Training of operational risk management model
- Application of the model
- Management actions

Each of the listed phases contains one or more steps, representing the specific activities that are necessary for its successful implementation. These steps are described in detail below.

Step 1: Define target variable and desired values

Classical management algorithms usually begin with defining the context and understanding the management problem to be solved. This is often done through an informal or semi-formal presentation of the structure of the problem, a process led by human experts. On the one hand, this makes significant sense, as it allows for the involvement of a wide group of stakeholders and allows for clear communication between them. On the other hand, such an approach with verbal definition of the problem significantly complicates the automation of its solution, as the presentation in the form of natural speech does not exclude logical gaps and does not guarantee a full scope of the space of the problem. This complicates its direct programming in an automated information system and in this sense does not provide a final criterion for determining the problem or an indicator for solving it. Instead, we propose to define one or more target variables and their desired values based on the business context and organizational needs.

Step 2: Powering real-time data by integrating with APIs

The second step is the beginning of the information support of the operational risk management process. The main task here is to find the necessary data for the process, as they include both the defined target variables in the first step and all available related data. The idea here is that the target variables often depend on a large number of different engines, which determine their dynamics and define our expectations for the realization of their values. If the projected realization differs significantly from the real one, this would mean the materialization of operational risk. To this end, a relatively accurate forecasting model should be defined, requiring relevant information. In this sense, it is important that the risk management module receives an automatic flow of data corresponding to the problem to be solved.

Step 3: Study the structure of the data and, if necessary, reduce their dimensionality

Having a large array of data has both its advantages and some disadvantages in their practical use. Among the latter, it is worth noting the need for significant space for their storage and high requirements for the computing power of the system when evaluating complex algorithms on large arrays of information. Therefore, it is possible and desirable to carry out an exploratory phase before the evaluation of the algorithm, which takes into account the basic structure of the data. At this stage, it is appropriate to display visualizations of the data and the relationships between them. Although this is not strictly necessary for the automated system, it is extremely useful for monitoring and control by human experts. The visualizations will help to better understand the process itself and organizational risks, as well as to more precisely control the results of the automated system.

Step 4: Iterative search for an optimal algorithm for risk modeling given data and constraints

In the next step of the application of the general algorithm, different models are trained to optimally describe the target variable set at the beginning. The types of statistical models (algorithms) that are tested depend on a wide range of characteristics, but of particular importance are the type of target variable, the characteristics of the analyzed data and the expected results. Two main approaches can be applied here. First, we can derive a set of popular algorithms with good results in previous research and similar applications and let the algorithm choose the optimal one among them. Secondly, we can get rid of this artificial constraint and instead search among a wide range of tens (hundreds) of potential algorithms, and finally bring out the optimal one among them.

Step 5: Evaluation of model parameters

After choosing the optimal model comes the step of calculating its parameters. Here again, it can be approached in two different ways. The first is to use the already calculated and tested coefficients of the optimal model from the previous step. The second is to evaluate the model on a different sample of data. There is some reason to approach the second way by expanding the volume of analyzed information, as it is likely that the larger sample size will lead to more accurate estimates of the model parameters. Additionally, in the fourth step, relatively smaller samples may be downloaded from the available data for the calculation of a large number of alternatives in terms of saving time and computational resources. This is no longer necessary here, as only one model is evaluated in the fifth step.

Step 6: Testing and classification of new observations according to a trained model

The application of the model itself on real instances of a given process is carried out in the sixth step. Here, the already evaluated model is used as the new observations, which power the control module through application program interfaces, are tested against it. Each of these observations is classified by assessing the extent to which the numerical values of its target variables deviate from the expected (predicted by the model) normal values. If the deviation exceeds the limits set by the anomaly criterion, it is classified as unusual and subject to additional action. This is essentially the step in deciding on the risk exposure. We emphasize that, unlike traditional approaches, the proposed algorithm allows to perform risk management at a very high level of detail (granularity), working at the level of atomic action or transactions, instead of at the level of business process or field of activity.

Step 7: Take action

Following the classification of an observation as legitimate or abnormal, specific actions should be taken to manage the risks associated with it. In traditional methodologies, human experts derive risk management strategies and activities and appoint a person responsible for each type of risk that carries them out. With the automation of the process and the transition to high granularity of management, this is no longer realistic - there is hardly an expert or even a department that can address hundreds of thousands or even millions of transactions every day. This requires the taking of automatic actions within the information system and other systems integrated with it. Here is the main question that dominates most discussions about the scope of digital transformation - to what extent actions should be taken automatically and to what extent the human workforce should be responsible for them. Achieving an optimal balance between the two is not a trivial task and it is appropriate to approach it formally by assessing the impact of the application of artificial intelligence in an economic context. Here we offer two main approaches that are applicable in modern organizations.

7.1. Notification of the person responsible for the risk

The first approach relies mainly on the role of the human workforce, which performs or at least explicitly authorizes the necessary actions. This follows a normal business process for risk management, in which a risk manager is appointed, who is expected to perform all its management activities.

7.2. Integration with other automatic action systems

The second approach provides full automation of operational risk management actions. This is achieved through integration with the transaction systems of the organization and implies clearly defined policies for this purpose. In this case, the operational risk management module is connected to the relevant information systems through a specific application programming interface and as a result of the risk classification can send requests for action to these systems.

Step 8: Iterative process of continuous improvement

The general algorithm for automated management of operational risks is essentially flexible and iterative, as it is possible to interact and intertwine different steps (eg the preparation of data for a specific algorithm can be done in the calculation of the model, not in the previous step). To enhance this iterativeness and ensure sustainable improvement of the algorithm, it also includes a process for continuous improvement. The process is formal, taking place both automatically at a time interval set by the analyst or user (eg once a month) and as a result of human intervention to start it (in case of unexpected changes in circumstances). This process involves reviewing the target variables, the current data structure, generating alternative models, selecting and calculating a new optimal one, without the need for all these steps to be affected. In this sense, only certain steps can be updated, and this most often involves recalculating the parameters of the model used with current data.

The general algorithm for automated management of operational risks aims to set the framework in which this process can be successfully performed, as the vast majority of its activities are performed by automated information systems.

Chapter 3: New Approaches to Managing Operational Risk

The classification of various observations as risky or non-risky ones is the most intuitive and common, but at the same time the best studied task in the field of risk management. From the point of view of the data used, it assumes that we have marked data, where the observations are divided into well-defined discrete classes (normal / anomalous). Hence, the target variable is naturally the variable that describes these classes. On the other hand, in many applications, the outputs of economic interest are not discrete but continuous variables. Examples include variables such as economic growth, probability of bankruptcy, return on an asset, income from certain activities or business lines, labor productivity and the like. Continuous output process control also involves the use of other types of algorithms in the field of statistics and machine learning, which are known as regression algorithms (Hastie et al., 2005).

The chapter aims to examine the automated application of optimal risk management algorithms in cases of long-term values of the target variable, to outline criteria for classifying certain observations as risky and to demonstrate the possibility of transferring these activities from human experts to an automated agent. To achieve these goals, five main situations of operational risks are considered, and the general algorithm for automated management is attached to them. These situations include modeling absences from the workplace, the popularity of the organization's online communication, managing incidents in the maintenance process, assessing the organization's real estate, and assessing adverse changes in the external environment. To identify and assess the risks in each of the situations we use a wide range of regression algorithms, and based on the degree of anomaly we identify potential risk observations. Here we show how on the basis of widely accepted statistical criteria for the quality of models and forecasts, as well as the level of anomaly of the observation we can conduct the overall process of risk identification and assessment in a completely autonomous way and without human intervention. This is key so that it can be executed automatically by an information system. At the same time, we recognize the importance of human control over the process, which is why as important steps we include the visualization and reporting of key relationships and results generated within the management process.

The significant number of current research in the field of statistical methods and machine learning also implies a rapid development of the methods, approaches and algorithms that are available for research purposes. In order to provide a comprehensive overview and optimal choice of a wide range of alternative algorithms, we use data to evaluate and test 136 of the most popular methods. The number of observations for training and testing of each of the algorithms is presented in an appendix, but the general principle is that in the comprehensive testing of classification algorithms we choose a random sample of the given information array, of which 80% is a training sample and the other

20% - a test sample. When testing, we calculate the area under the ROC-curve of each algorithm, but also a measure of its complexity. As an approximation of the complexity measure, we use the time required for the calculation by standardizing the longest required time to 100% and presenting the remaining times as a part (proportion) of it. In this sense, the complexity measure varies in the range from 0% to 100%. The computational time of an algorithm is highly dependent on the infrastructure and method of calculation used, and it is of particular importance whether the calculation is distributed or not. It is misleading to report the "clean" time required, as it will depend on the machines or clusters of machines used. The complexity measure partially solves this problem by deriving time relations rather than absolute values. Although some problems remain - for example, the computational time would change from the type of processor, architecture, load management, etc., the measure of complexity is a satisfactory approximation of how resource-intensive an algorithm is.

Conducting a direct marketing campaign

The forecast accuracy of the considered methods is summarized in the histogram of Figure 3. First of all, the significant variation in the accuracy between the different algorithms used is impressive. Even if we ignore the extreme values (positive and negative), the main part of the distribution changes in the range from 0.5 to 0.78. This suggests that some algorithms are significantly more suitable for a certain type of task than others. In this sense, the choice of an appropriate classification algorithm can lead to a very significant difference in the generated results and hence - in the created value of the business. Secondly, we note that the distribution shown is close to normal, observing a peak around AUROC = 0.7. This would be the expected predictive accuracy of the "average" algorithm solving this task. Third, we consider the relatively high number of algorithms that do not add value (area under the curve of 0.5), emphasizing that they should be avoided.

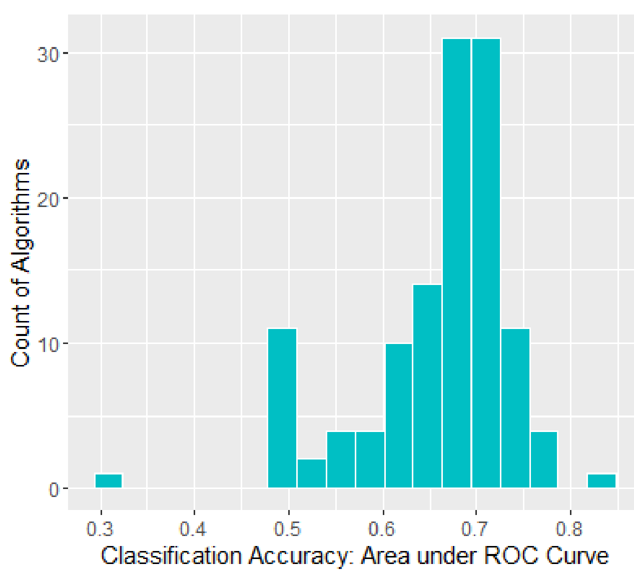

Figure 3: Histogram of Classifier Accuracy in Portuguese Marketing Campaign Data

The first ten classification methods with the best results shown are presented in Table 3: Top 10 classification methods with the highest forecast accuracy. It is noteworthy that different models of discriminant analysis perform extremely well in this task, and the best classifier is the robust regularized discriminant analysis (rrlda) with an area below the performance curve of 0.82. Solution trees and different types of ensemble algorithms are also among the top ten classifiers, with the area under the ROC curve for algorithms in places 2 to 10 varying in the range of 0.73 to 0.77. We also emphasize that the most accurate classification is not achieved by the most resource-intensive algorithms, as the Top 10 includes optimized and relatively fast algorithms.

Table 3: Top 10 classification methods with the highest forecast accuracy

Algorithm Type	Implementation Method	Area under ROC Curve	Complexity Measure
Robust Regularized Linear Discriminant Analysis	rrlda	0.824	1.8%
Soft Independent Modeling of Class Analogies, SIMCA	CSimca	0.773	0.3%
Rule-Based Classifier	JRip	0.767	7.8%
Mixture Discriminant Analysis	mda	0.764	0.1%
Conditional Inference Tree	ctree	0.762	0.3%
C4.5-like Trees	J48	0.754	3.0%
Model Averaged Neural Network	avNNet	0.740	3.9%
ROC-Based Classifier	rocc	0.738	0.7%
Bagged AdaBoost	AdaBag	0.734	2.3%
Tree-Based Ensembles	nodeHarvest	0.734	56.7%

Conducting credit card operations

Servicing credit card balances due is a key problem in the financial sector, and the inability to do so and potential fraud can have a significant effect on the financial flows and solvency of subsidiaries. In this context, it is particularly important to choose the optimal algorithm, as even small improvements in forecasting accuracy can lead to the unlocking of significant value for creditors. For this purpose we perform comprehensive testing of 136 basic algorithms in the field of machine learning and analyze their accuracy in classification.

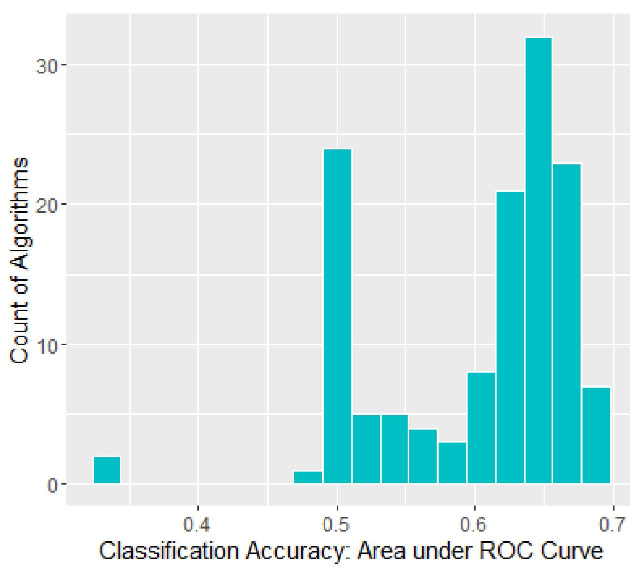

Figure 4: Histogram of Classifier Accuracy in the Taiwan Credit Card Debt Data

Figure 4 summarizes the data for the forecast accuracy of the alternative algorithms, measured as the area under the performance curve (ROC-curve). A large number of the considered algorithms have an AUROC of about 0.50, which is a result equal to that of the chance - therefore this first peak of the distribution shows the fruitless algorithms for this task. In the histogram we observe a second peak with values around 0.65, as the vast majority of the considered algorithms are concentrated in the range from 0.62 to 0.68. The best algorithms tend to an area of 0.7, but in reality none exceeds this limit. We consider the classification task on credit card data as a relatively difficult one, which explains the results obtained.

The ten best classification methods are presented in Table 3. It is noteworthy that the group is dominated by two main types of models - that of machines with supporting vectors and classification and regression trees (CART). The highest result is obtained by the machine of supporting vectors with polynomial kernel, calculated by the method of least squares with AUROC = 0.684, followed by decision trees of type C5.0 and three other variations of machines with supporting vectors (all with AUROC - 0.682). In the top ten are two more tree-based methods, one ROC-based classifier and an ensemble adaptive enhancement model. We note that the accuracy of all these algorithms is very similar and in practice there will be relatively small differences that would only matter when processing large data sets. As for the time required for calculation, the best algorithms are again not the most resource-intensive. The optimal method is nearly a hundred times faster than the slowest, the second best is a thousand times faster, and the third is 167 times faster. This shows that in this task we again see an opportunity to balance between the computational load and the accuracy of the results.

Table 4: Top 10 classification methods with the highest forecast accuracy

Algorithm Type	**Implementation Method**	**Area under ROC Curve**	**Complexity Measure**
Least Squares Support Vector Machine with Polynomial Kernel	svmPoly	0.684	1.1%
Single C5.0 Ruleset	C5.0Rules	0.682	0.1%
SVM Linear Weighted	svmLinear Weights	0.682	0.6%
SVM Linear	svmLinear	0.682	0.2%
SVM Linear2	svmLinear2	0.682	0.2%
ROC-Based Classifier	rocc	0.680	0.2%
CART	rpart1SE	0.679	0.1%
Bagged AdaBoost	AdaBag	0.677	37.4%
Boosted Tree	bstTree	0.677	1.9%
Boosted Classification Trees	ada	0.673	3.2%

Granting Loans

Credit risk modeling is a classic classification task and standard machine learning algorithms can be applied to it. For this purpose, we evaluate 136 of them, and the summarized results for their forecast accuracy are presented in Figure 5.

The distribution we observe in this case differs significantly from the normal one. We observe a peak of the algorithms with forecast accuracy around the chance (AUROC = 0.5), followed by a relatively uniform distribution of algorithms with forecast accuracy in the range 0.52 to 0.64. Most of the considered algorithms have an accuracy in the range of 0.68 to 0.70, which can be said to be our expectations for the "average" algorithm suitable for this particular task. A small number of algorithms with AUROC> 0.70 are observed, which are also the best performing classifiers for the target variable of the considered data set. The first ten classifiers with the highest forecast accuracy are presented in Table 5.

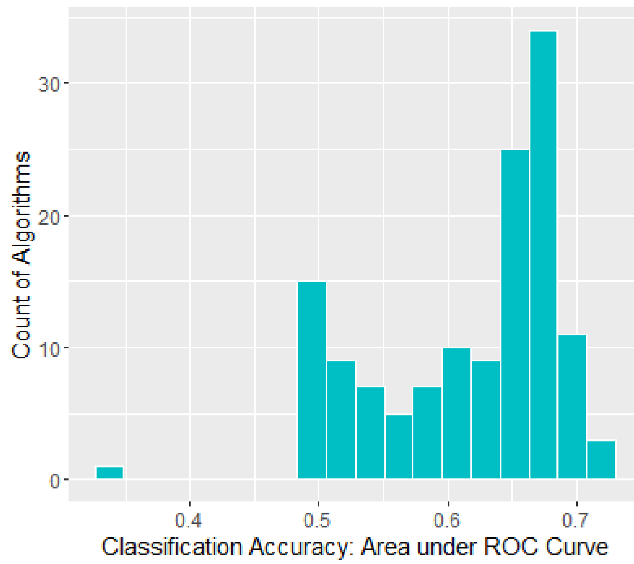

Figure 5: Histogram of Classifier Accuracy in the German Credit Data

They all have an area under the ROC curve of at least 0.7, with the best of them - the regularized random forest - reaching a value of 0.73. It is noteworthy that this group is dominated by different types and methods for calculating random forests, as they make up half of the ten best algorithms. In addition, there are methods for amplifying the gradient, a specific type of neural network (multilayer perceptron), as well as a variant of discriminant analysis - localized linear discriminant analysis. Again, the most computationally difficult algorithms do not give the best results. The most accurate classifier is calculated 2.6 times faster than the most resource-intensive, and the second most accurate - 13.5 times faster.

Table 5: Top 10 classification methods with the highest forecast accuracy

Algorithm Type	**Implementation Method**	**Area under ROC Curve**	**Complexity Measure**
Regularized Random Forest	RRF	0.730	38.5%
eXtreme Gradient Boosting	xgbLinear	0.718	7.4%
Regularized Random Forest	RRFglobal	0.712	5.2%
eXtreme Gradient Boosting	xgbDART	0.707	26.3%
Multi-Step Adaptive MCP-Net	msaenet	0.702	5.1%
Random Ferns	rFerns	0.701	2.0%
Localized Linear Discriminant Analysis	loclda	0.701	0.8%
Random Forest	rf	0.700	1.7%
Random Forest	ranger	0.698	1.6%
Gradient Boosting Machine	gbm	0.698	0.2%

Estimating the Probability of Default of External Partners

The task of determining the class affiliation of companies to whether or not they will go bankrupt is proving to be of considerable difficulty. Among the seven most popular approaches, only the random forest shows satisfactory results. These are the cases in which the search for an optimal algorithm should be significantly expanded, as here we test 136 different alternatives. The respective models were calculated on a training sample, and their predictions were tested on a test sample. The distribution of their accuracy, measured by the area under the performance curve, is summarized in the histogram of Figure 6. A huge number (over 40) of the evaluated algorithms have a forecast accuracy around AUROC = 0.5, which is exactly equal to the classification relative to the unconditional probability of belonging to a given class. The difficulty of the problem is also underlined by the fact that there are some algorithms with an area below the ROC curve of less than 0.5, which is a result worse than the randomly generated one. We observe a slight peak of the forecast accuracy at values of the area of 0.6, as the best classification algorithms reach AUROC forecast accuracy above 0.7. We note significant differences between the results of different methods, however, there are a minority of approaches with a fairly high forecast accuracy.

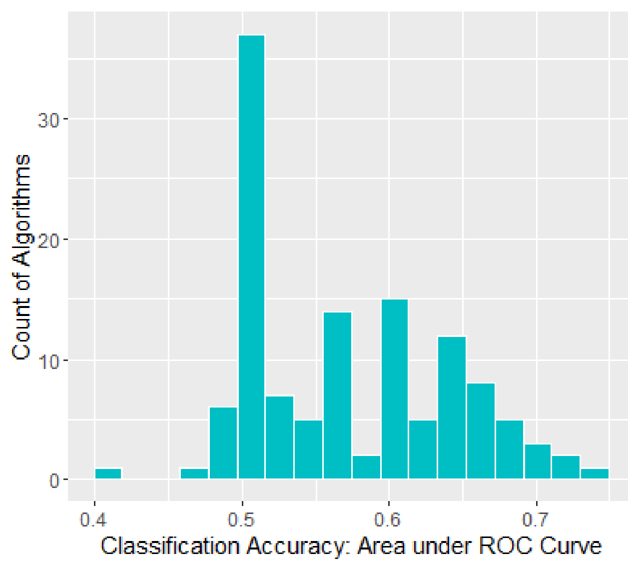

Figure 6: Histogram of Classifier Accuracy for Polish Company Defaults Data

The most accurate ten algorithms are presented in Table 6. The most optimal among them is the robust soft independent modeling of class analogies, RSIMCA. This method is relatively less well known and used in economics and business, but essentially involves supervised analysis, which divides the data into main components and constructs subspaces based on those components, which are then used for classification. . For more details, we direct the reader to the original development of Brandon & Hubert (2005), as well as to the study of Fauziyah et al. (2018). The RSIMCA model has an area under the ROC curve of 0.733 and is more than three hundred times faster than the slowest algorithm - the Method for slow derivation of rules, which is second in forecast accuracy with 0.721. In third place with very close forecast accuracy (AUROC - 0.714) are random trees, followed by six other methods from the family of decision trees or random forests. All of them are relatively fast and in this sense require relatively more limited computing resources.

Table 6: Top 10 classification methods with the highest forecast accuracy

Algorithm Type	Implementation Method	Area under ROC Curve	Complexity Measure
Robust SIMCA	RSimca	0.733	0.3%
Patient Rule Induction Method	PRIM	0.721	100.0%
Random Ferns	rFerns	0.714	1.7%
CART	rpart1SE	0.693	0.1%
CART	rpart2	0.693	0.0%
Single C5.0 Ruleset	C5.0Rules	0.693	0.2%
Rule-Based Classifier	PART	0.679	0.3%
Regularized Random Forest	RRF	0.676	35.6%
Shrinkage Discriminant Analysis	sda	0.676	0.2%
Bagged AdaBoost	AdaBag	0.674	0.9%

The tenth place in terms of forecast accuracy is ranked by a specific method for discriminant analysis, again registering a relatively high accuracy - AUROC = 0.676. We recognize that it is in tasks with higher difficulty that it makes a lot of sense to test a wide range of alternatives and to choose the optimal one for the given situation and type of data. This allows the risk management process to take advantage of algorithms that are particularly suitable for specific niche situations, but on average statistically have a poorer performance in a wider range of applications and are therefore not popular and often used.

E - commerce activities

To find the optimal classification algorithm in the study of online consumer behavior, we evaluate 136 different machine learning algorithms. The summarized results for their forecast accuracy, measured by the area under the performance curve, are shown by the histogram of Figure 7.

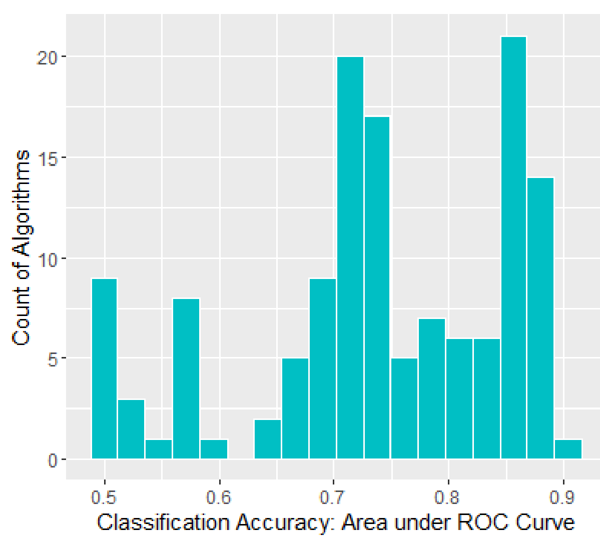

Figure 7: Histogram of Classifier Accuracy for Online Purchases Data

The average forecast accuracy in the classification of online behavior is significantly higher than the other considered risk situations. It is noteworthy that the distribution of areas is characterized by two peaks - one is around AUROC = 0.70-0.75, and the other - at values in the range 0.85-0.90. The best classification algorithm scores even above 0.90. We also take into account the significant variance in the results of the calculated methods. Many of them have results close to the chance, but a significant minority register a very high forecast accuracy. This emphasizes the importance of choosing the optimal algorithm due to the very different quality of the generated forecasts.

Table 7: Top 10 classification methods with the highest forecast accuracy

Algorithm Type	**Implementation Method**	**Area under ROC Curve**	**Complexity Measure**
Rotation Forest	rotationForestCp	0.902	1.0%
Weighted Subspace Random Forest	wsrf	0.889	4.4%
Patient Rule Induction Method	PRIM	0.886	14.3%
Boosted Tree	bstTree	0.885	1.4%
CART	rpart	0.880	0.0%
CART or Ordinal Responses	rpartScore	0.880	1.4%
Conditional Inference Tree	ctree2	0.880	0.1%
C5.0	C5.0	0.879	0.3%
Cost-Sensitive C5.0	C5.0Cost	0.879	0.7%
DeepBoost	deepboost	0.876	4.9%

The most accurate ten algorithms according to the area under the ROC curve are presented in Table 7. The rotary forest has the best performance with AUROC = 0.902, followed by the random forest with weighted spaces (0.889), the method for slow derivation of rules PRIM (0.886) , an enhanced decision tree (0.885), and a series of methods from the family of classification and regression trees (all with 0.880). We consider the family of methods of random forest, as well as the trees that make them up, as the most optimal approach for solving the considered problem of consumer behavior in a digital environment. From the point of view of the time and resources required to calculate these methods, we emphasize that again the most resource-intensive methods do not lead to the best forecast results. On the contrary, the optimal algorithm is calculated 100 times faster than the slowest one, and we observe similar and better ratios in the subsequent methods in the top ten. This sustainable result in all considered tasks shows the possibility for simultaneous optimization of both the accuracy and the used IT resources and computing infrastructure.

The review of a wide range of popular algorithms in the field of machine learning on five different operational situations allows us to outline some basic conclusions. First of all, the best algorithms for each of the considered problems do not overlap - in each of the individual classifications we observe differences in the best algorithms for its solution. This is probably due to the fact that different families of algorithms and their specific methods are they handle certain types of data better, but generate worse results with other types. This is the well-known theorem for the lack of free lunch in optimization (Branden & Hubert, 2005) and emphasizes that it is a theoretically and practically unsuccessful approach to use the same methods for each task.

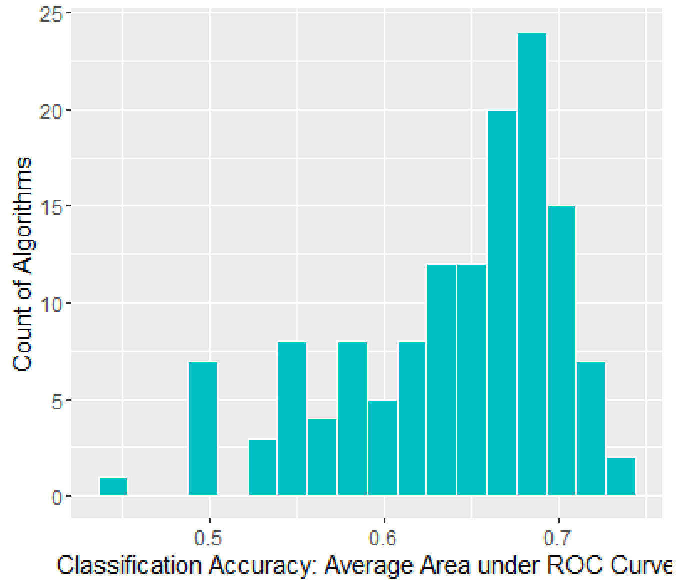

Figure 8: Histogram of Averaged Classifier Accuracy Across All Datasets

Secondly, we note that the forecast accuracy between the different algorithms can vary in extremely wide intervals. Figure 8 summarizes the distribution of the average values of the area under the performance curve for all considered algorithms. The average forecast accuracy varies in the range from AUROC = 0.50 to 0.74. This emphasizes that the importance of choosing the optimal algorithm is not only theoretically justified, but can have significant practical consequences. This result further underscores the importance of a comprehensive search for the best operational risk management algorithms, as improvements in forecasting accuracy have the potential to generate huge business value at the appropriate scale of business operations.

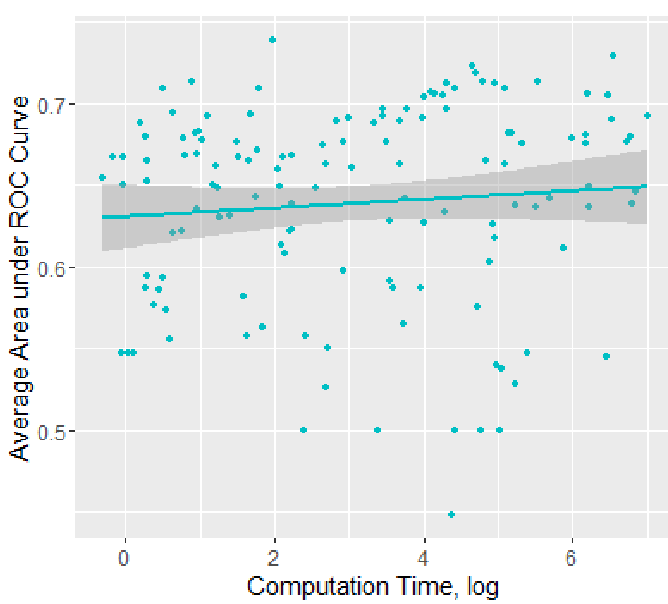

Figure 9: Relationship between Classification Accuracy and Algorithm Computation Time (log.)

Third, we note that certain families of algorithms tend to show better results than others. In particular, the different methods of random forest are often among the best algorithms for solving each of the considered problems. In the analysis, we noticed the usual tendency for them to adapt to certain analyzed data, but nevertheless they show excellent results in new test samples. From the point of view of classical classification algorithms, it seems that discriminant analysis in its various variants has quite good forecast results. As these methods are in most cases highly optimized, they could be a reasonable compromise in situations where a significant amount of data needs to be analyzed under conditions of limited computational resources.

Fourthly, we emphasize that the most computationally resource-intensive algorithms do not necessarily reach the most accurate forecast results. In each of the considered problems the best method for classification is not the one that takes the most resources for its calculation. Among the first ten methods according to the area under the performance curve for each of the tasks are methods that are tens and even hundreds of times faster than the slowest one. Figure 9 graphically presents this relationship between the mean area under the ROC curve for all methods on all tasks and the logarithm of the time required to estimate them. Graphically, we report a weak positive relationship between the two, but in its study within a linear regression model, this relationship does not reach statistical significance ($p = 0.323$). In this sense, it is possible to choose the optimal ratio between forecasting accuracy and the required computational resources for a given method, so as to generate maximum benefit for the risk management process.

The results shown in this chapter demonstrate the possibility of automating important steps in the process of identifying and assessing operational risks. The automatic selection of a risk assessment algorithm based on a single indicator (area under the performance curve) and its application allow the management process to take place at the level of individual monitoring or event and release employees from the unbearable responsibility to monitor and assess in real time all events relevant to the process. Moreover, the use of supervised learning algorithms in human experts remains the important task of determining target variables, their desired values and monitoring and control over the operation of the automated information system. This allows simultaneous optimization of a large part of the necessary activities and preservation of the valuable human expert element in quality calibration of the used models and systems.

Chapter 4: Automated Regression Algorithms for Operational Risk Management

The classification of various observations as risky or non-risky ones is the most intuitive and common, but at the same time the best studied task in the field of risk management. From the point of view of the data used, it assumes that we have marked data, where the observations are divided into well-defined discrete classes (normal / anomalous). Hence, the target variable is naturally the variable that describes these classes. On the other hand, in many applications, the outputs of economic interest are not discrete but continuous variables. Examples include variables such as economic growth, probability of bankruptcy, return on an asset, income from certain activities or business lines, labor productivity and the like. Continuous output process control also involves the use of other types of algorithms in the field of statistics and machine learning, which are known as regression algorithms (Hastie et al., 2005).

The purpose of their use is to determine with maximum accuracy the expectation for realization of the target variable and if it is likely to go beyond the desired or acceptable limits (operational risk event) to take management measures to reduce the negative effect.
This chapter aims to examine the automated application of optimal risk management algorithms in the case of long-term values of the target variable, to outline criteria for classifying certain observations as risky and to demonstrate the possibility of transferring these activities from human experts to an automated agent. . To achieve these goals, five main situations of operational risks are considered, and the general algorithm for automated management is attached to them. These situations include modeling absences from the workplace, the popularity of the organization's online communication, managing incidents in the maintenance process, assessing the organization's real estate, and assessing adverse changes in the external environment. To identify and assess the risks in each of the situations we use a wide range of regression algorithms, and based on the degree of anomaly we identify potential risk observations.

Excessive absences from work
Given the significant choice of alternative regression algorithms, we should consider the possibility that some of them have significantly better forecast capabilities than others. To verify this, we evaluate 106 alternative algorithms on data for excessive absence from the workplace, presenting their forecast accuracy, measured by the root mean square error, in Figure 10.

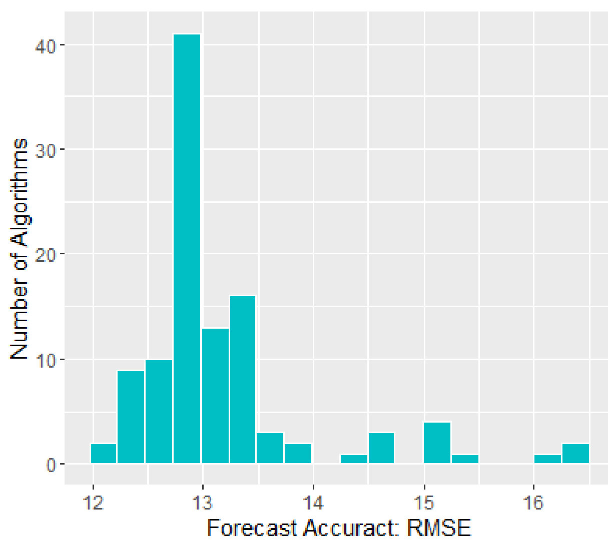

Figure 10: Histogram of forecast accuracy for workplace absenteeism data

On it we note a significant group of algorithms with forecast accuracy in the vicinity of RMSE = 13, as almost all considered alternatives have a root of the root mean square error for this task in the range from RMSE = 12 to RMSE = 14. The best algorithms register RMSE ~ 12, but there are several with particularly poor performance, in which RMSE> 15. Considering the results of this problem, we note that almost all algorithms give satisfactory results, and the difference between the best and average ones is relatively small. The key in this case is to avoid the use of any of the particularly weak methods, as their results register significantly higher error rates than the average.

Table 8: Forecast accuracy of the top ten algorithms for workplace absenteeism data

Algorithm	**Method**	**Mean Error, ME**	**Root Mean Squared Error, RMSE**	**Mean Absolute Error, MAE**	**Complexity Measure**
CART	rpart2	0.621	12.172	5.376	0.1%
eXtreme Gradient Boosting	xgbDART	0.043	12.174	5.459	1.8%
Random Forest by Randomization	extraTrees	-0.314	12.229	5.418	2.9%
Gaussian Process with Radial Basis Function Kernel	gausspr Radial	0.054	12.287	5.573	0.7%
Bagged CART	treebag	0.205	12.325	5.459	0.4%
Regularized Random Forest	RRFglobal	-0.603	12.330	5.724	0.4%
Conditional Inference Random Forest	cforest	-0.141	12.371	5.581	14.5%
Conditional Inference Tree	ctree	0.798	12.381	5.387	0.3%
Regularized Random Forest	RRF	-0.587	12.424	5.792	2.9%
Random Forest	rf	-0.500	12.430	5.688	8.5%

The methods with the lowest forecast errors are presented in Table 8. We immediately notice that six of the top ten methods are different variants of the random forest family. The best forecast accuracy has one of the applied methods of classification and regression trees (rpart) with the root of the mean square error 12.17 and the mean error of 0.62, followed by the algorithm for extreme gradient amplification with RMSE = 12.17 and ME = 0.04, respectively.

Figure 11: High Risk Anomalous Observations According to Algorithms Consensus

We take into account the extremely small difference in forecast accuracy between the ten best methods - it varies from RMSE = 12.17 to RMSE = 12.43, which emphasizes the possibility of choosing a method among this already narrow set to be made on the basis of other considerations - necessary computational resources, organizational experience and consideration, possibility for interpretation of results and integration with other systems, etc. From the point of view of the resources used, we note that none of the top ten algorithms is particularly demanding from this point of view. The best method is 1,000 times faster than the slowest alternative and has a calculation time comparable to that of multidimensional linear regression. These results are partly due to the popularity of random forests, which lead to a number of applied methods within this family, and they often seek to optimize the algorithm in terms of resources used and hence the calculation time. Looking at the hours of absence from work in relation to the distance to the home of employees, we do not notice a clear trend among the anomalous observations. Here again, we note that extreme observations have been successfully identified using the proposed criteria, and the more extreme an observation is, the more likely it is to be defined as abnormal according to more than one criterion. Figure 11 focuses on the most risky observations in the sample according to two or more criteria. It allows to prioritize the various events representing potential risks, addressing them from the highest priority (anomalies according to at least 3 algorithms) and move on to lower priority ones. This approach ensures the rational use of organizational resources and their focus on activities that would unlock the potential greatest business value.

Online communication

In search of an optimal forecast method for the risks of online communication, we calculate the forecast for the target variable using over 100 alternative algorithms, and the distribution of their predictive accuracy (RMSE) is presented graphically in Figure 12. The vast majority of tested algorithms have very good forecast accuracy from RMSE <11,000. We observe a strongly shifted distribution, in which a large number of approaches generate close to optimal results, and there are a smaller number of algorithms with much lower performance.

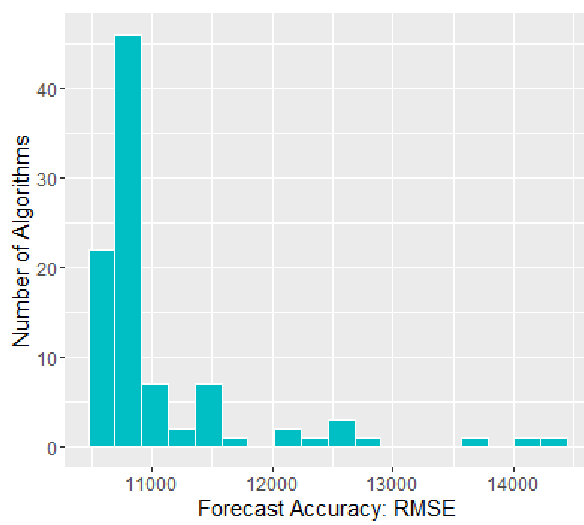

Figure 12: Histogram of forecast accuracy for online news sharing data

The ten approaches with the highest forecast accuracy are presented in Table 9. It is noteworthy that six of them represent different methods for calculating a random forest. The best algorithm in this case is ranger, which is a highly optimized application of the random forest, which achieves comparable forecasting accuracy at a very low cost of computational resources. Its root mean square error is only RMSE = 10514, and the other ten algorithms are in this neighborhood. Apart from the methods for calculating a random forest, we also notice algorithms of Gaussian process, Bayesian and lasso regression, as well as a generalized linear model. From the point of view of all calculated measures for forecast accuracy, the presented top ten algorithms are difficult to distinguish and the choice for the optimal one can be made on the basis of an appropriate resource or organizational criterion.

Table 9: Forecast accuracy of the top ten algorithms for online news sharing data

Algorithm	**Method**	**Mean Error, ME**	**Root Mean Squared Error, RMSE**	**Mean Absolute Error, MAE**	**Complexit y Measure**
Random Forest	ranger	384.52	10514.60	3566.87	0.35%
Regularized Random Forest	RRF	247.73	10532.25	3630.94	9.92%
Regularized Random Forest	RRFglobal	231.90	10533.87	3633.60	1.46%
Random Forest	rf	213.69	10573.80	3662.55	0.52%
Random Forest by Randomization	extraTrees	209.17	10584.73	3639.87	3.33%
Gaussian Process with Radial Basis Function Kernel	gaussprPoly	311.21	10636.78	3703.16	1.42%
Bayesian Ridge Regression	bridge	349.77	10651.67	3637.69	0.30%
Conditional Inference Tree	cforest	366.84	10656.54	3646.23	0.68%
glmnet	glmnet	370.03	10657.97	3607.08	0.00%
Lasso Regression	lasso	369.31	10659.49	3608.71	0.01%

For this purpose, the measure of complexity can be used as an approximation of the resource needs of the given algorithm. With the exception of the regulated random forest, all other algorithms are relatively fast, including the ranger being about 300 times faster than the slowest option considered.

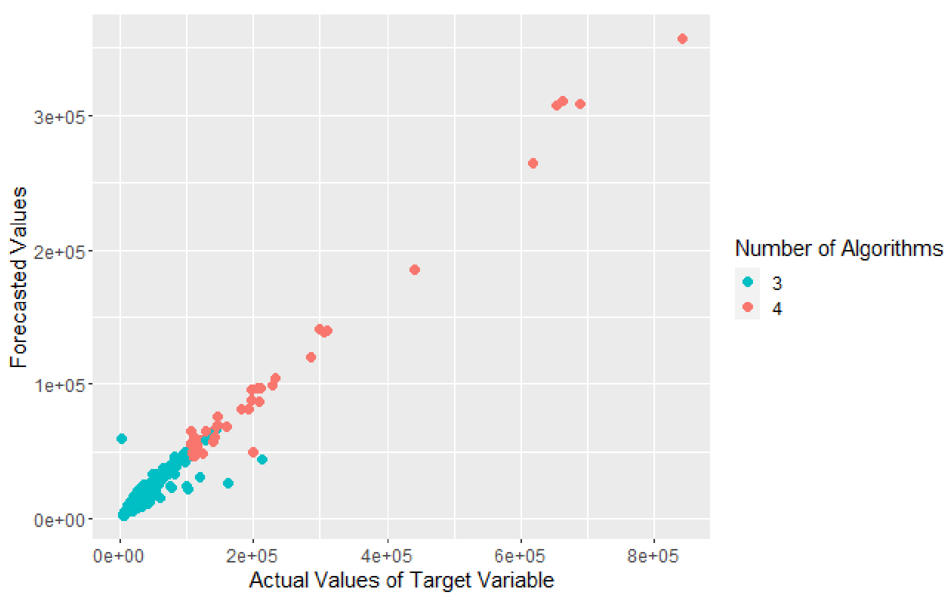

Figure 13: High Risk Anomalous Observations According to Algorithms Consensus

In this sense, it is an appropriate choice for a regression algorithm for solving such problems. For the purposes of operational risk management, we propose to analyze observations that have been identified as abnormal by at least half of the algorithms. Applying this criterion significantly narrows the number of cases that are subject to active management - only 1,827 (4.6%) observations are classified as abnormal according to at least 3 criteria and 43 (0.1%) - according to four. These observations are presented visually in Figure 13 and are subject to risk management follow-up.

Valuation of asset prices (real estate)

On the basis of the real estate data, 106 alternative models from the field of machine self-learning were evaluated, and their forecast accuracy was studied in detail in relation to the forecast errors generated by them. A histogram of their predictive accuracy relative to the root mean square error is presented in Figure 14. It is noteworthy that the vast majority of methods have a deviation of RMSE in the range of 7 to about 9. The best algorithms among the studied have a predictive accuracy of RMSE <6.5, and those with the worst results can reach a value of RMSE above 25. It is noteworthy that although in general in solving this problem we notice a significant grouping around one value, it also has extremely poor results. A small number of algorithms also have slightly better performance than average accuracy. In this sense, there is potential business value in testing and choosing the best among them.

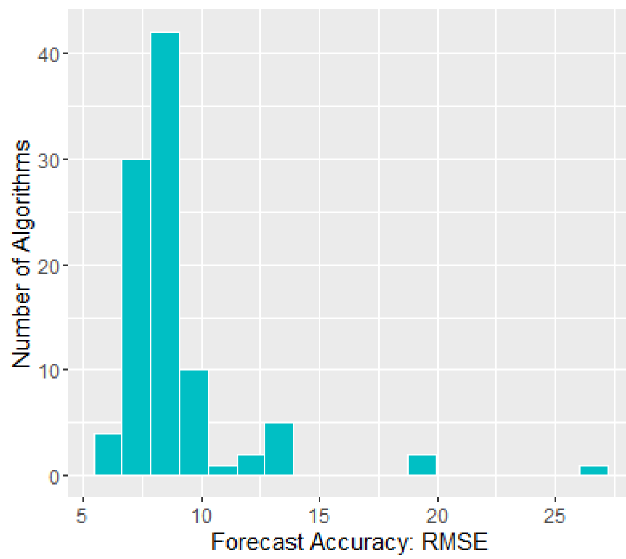

Figure 14: Histogram of forecast accuracy for housing prices data

The ten approaches with the lowest root mean square error are presented in Table 10. It is immediately striking that seven of them are different applications within the random forest family. All of them registered extremely good forecast accuracy, as the root of the root mean square error of the predictions compared to the test sample is in the range from RMSE = 6.46 to RMSE = 7.10. The other three non-random forest algorithms are 2 based on a kernel function and one based on a Gaussian process, with a forecast accuracy of about RMSE = 7.10.

The complexity measure, which takes into account a proportional calculation time to the most resource-intensive algorithm, also varies widely. The best method – that of the regularized random forest is only 35% faster than the slowest in the sample. On the other hand, the second best - the quantile random forest - is nearly 20 times faster than the most resource-intensive, and the difference in forecast accuracy between the two is almost imperceptible.

Table 10: Forecast accuracy of the top ten algorithms for housing prices data

Algorithm	**Method**	**Mean Error, ME**	**Root Mean Squared Error, RMSE**	**Mean Absolute Error, MAE**	**Complexity Measure**
Regularized Random Forest	RRF	-0.750	6.459	4.831	64.8%
Quantile Random Forest	qrf	0.001	6.470	4.695	5.2%
Regularized Random Forest	RRFglobal	-0.832	6.568	4.890	9.9%
Random Forest	ranger	-0.878	6.600	4.884	7.0%
Parallel Random Forest	parRF	-0.936	6.689	4.965	3.8%
Random Forest	ranger	-0.943	6.689	4.908	4.0%
Radial Basis Function Kernel Regularized Least Squares	krlsRadial	-0.435	7.068	5.388	14.8%
Bayesian Additive Regression Trees	bartMachine	-0.795	7.076	5.353	11.1%
Random Forest by Randomization	extraTrees	-0.948	7.081	5.137	7.8%
Gaussian Process with Polynomial Kernel	gaussprPoly	-0.501	7.082	5.443	2.1%

This emphasizes that even with this type of task it is possible to find the optimal point between the benefits and costs in the calculation of the given algorithms. Moreover, the speed of calculation of the algorithm shows the possibility of switching from asynchronous to synchronous operations, ie. from calculation of models and their subsequent use and updating in the future to real-time analytics, which is used and trained simultaneously. Figure 15 presents the observations with the potentially highest risk - they are classified as extreme compared to three or four of the four approaches used. The graph allows us to trace that these observations do not follow the expected trend and do not show a relationship between forecast and realization. In this sense, they are completely unexpected and represent a potential realization of operational risks.

Figure 15: High Risk Anomalous Observations According to Algorithms Consensus

Sharp changes in market demand

Based on the considered data, we calculate 109 alternative forecast models with different regression algorithms in the field of machine self-learning. Their forecast accuracy, measured by the root of the root mean square error, is presented in Figure 16. As can be seen from the histogram, the small number of observations leads to unstable forecast results. There are a large number of algorithms that, through oversupply, reach extremely low RMSE values close to zero. We also notice a small number of algorithms with extremely poor performance - RMSE in the vicinity of 300. This result is significantly weaker than a naive prediction for each value (eg the average for the sample).

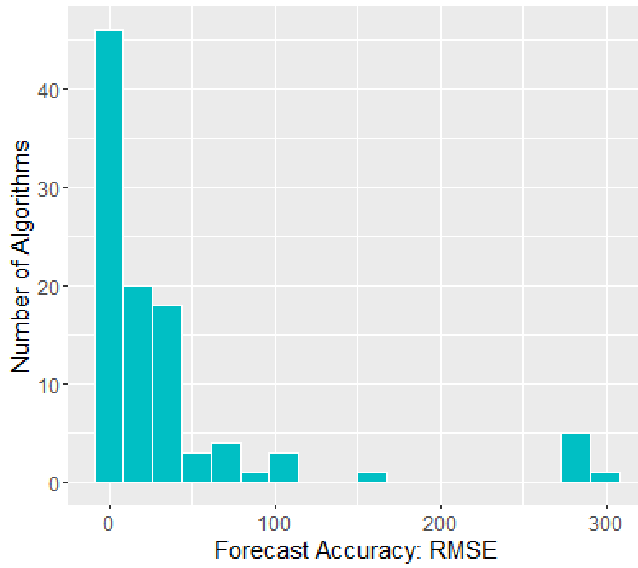

Figure 16: Histogram of forecast accuracy for logistics demand data

As a potential way to avoid relying on models with uncertain performance outside the training sample, it is appropriate not to consider models with RMSE <1, emphasizing that for the vast majority of them it is close to zero. Among the other algorithms we can highlight those with the highest forecast accuracy, as the ten best are presented in Table 11. The most optimal models are the ensemble multidimensional adaptive regression splines (MARS), as they reach RMSE = 2.51 and RMSE = 2.61 . In this task, a number of representatives of the family of generalized linear models, as well as machine models with supporting vectors, clearly stand out.

Again, we note that the best performing algorithms are not the most resource-intensive, which allows based on the forecast accuracy and the measure of complexity to make an optimization decision for the optimal model for solving the given task. Given the accuracy and measure of complexity of the model of ensemble multidimensional adaptive regression splines (MARS) without restrictions, we choose it as an algorithm for subsequent testing. We emphasize that with such a sample size, all results should be interpreted with caution and that in the practical application of the proposed method, ways should be sought to avoid samples of similar size if possible. Figure 17 examines in more detail only the observations that are classified as anomalous by at least one algorithm (criterion). The vast majority of them are unusual according to only one approach and this is most often the local factor of extreme value, while there is only one observation, which is anomalous according to 3 algorithms and one - compared to four. We emphasize that it is appropriate in situations of unsupervised self-learning as potentially risky to consider observations that are marked as such by more than one criterion.

Table 11: Forecast accuracy of the top ten algorithms for logistics demand data

Algorithm	**Method**	**Mean Error, ME**	**Root Mean Squared Error, RMSE**	**Mean Absolute Error, MAE**	**Complexity Measure**
Bagged MARS	bagEarth	-0.056	2.509	2.076	1.12%
Bagged MARS using gCV Pruning	bagEarthGCV	0.107	2.605	2.185	0.48%
glmnet	glmnet	-1.188	2.780	2.384	0.14%
Gaussian Process with Linear Kernel	gaussprLinear	-0.966	2.843	2.357	0.28%
Bayesian Regularized Neural Networks	brnn	-0.910	2.853	2.374	0.08%
Support Vector Machines with Linear Kernel	svmLinear2	-3.153	5.865	5.129	0.07%
Support Vector Machines with Linear Kernel	svmLinear	-3.153	5.865	5.129	0.28%
Gaussian Process with Polynomial Kernel	gaussprPoly	-3.527	7.345	5.660	0.16%
Least Squares Support Vector Machine with Polynomial Kernel	svmPoly	-3.146	7.660	6.936	0.29%
Boosted Generalized Linear Model	glmboost	-2.669	7.863	6.994	0.10%

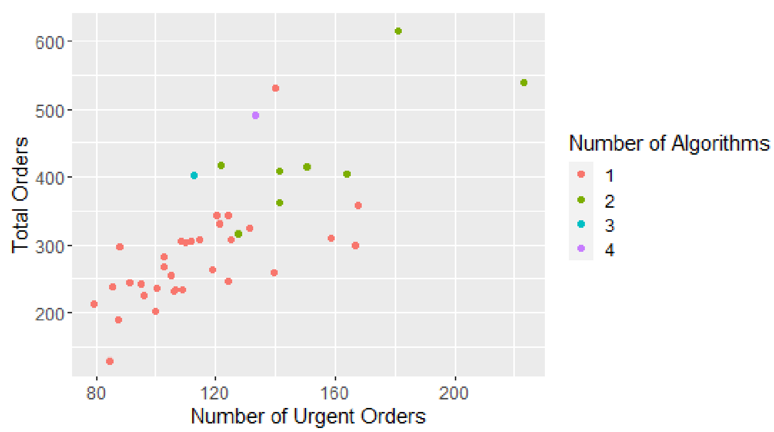

Figure 17: High Risk Anomalous Observations According to Algorithms Consensus

Support ticket processing

Using an automated approach to calculate the predicted values for signal processing time, we can significantly expand the range of tested algorithms for machine self-learning. Based on the current data, we evaluate 109 different algorithms, of which 102 reach convergence and can be used to generate forecasts. The distribution of their accuracy, measured by the root mean square error (RMSE) is shown graphically in the histogram of Figure 18. The vast majority of algorithms register RMSE about 450, noting the long queue of the distribution - it shows the presence of a large number of algorithms with special poor performance. Due to the specifics of the tasks and the data, it is possible to observe the tendency of certain algorithms to cope especially well with some tasks, which is compensated by a particularly poor performance of others. Therefore, it is unsuccessful to use the same algorithm for each task based only on past good results from solving other tasks with another data set. This is one of the reasons to include in the general algorithm for automated assessment and management of operational risks proposed here a component of testing and testing of a wide range of alternative algorithms.

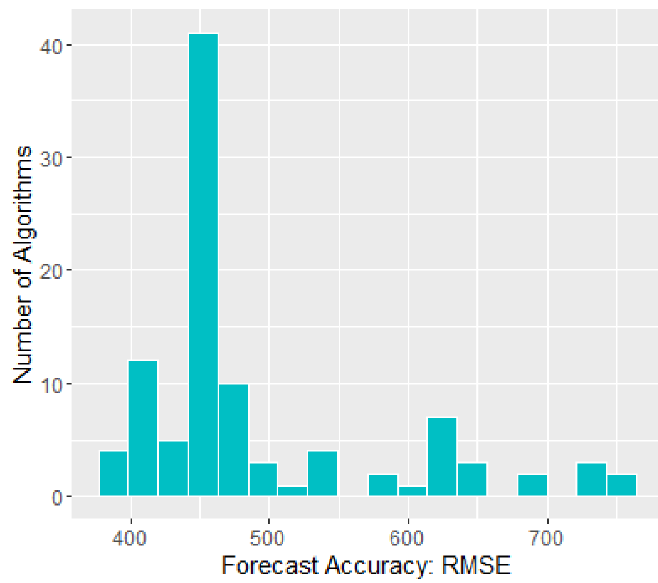

Figure 18: Histogram of forecast accuracy for support ticket processing data

The ten approaches with the highest forecast accuracy are summarized in Table 12. Eight of the ten methods are variations and different applications of the random forest algorithm. The highest forecast accuracy is achieved by the RRFglobal method, as its forecast accuracy is RMSE = 388.8, followed by five other different applications of random people with comparable error levels. The list of the ten most accurate algorithms also includes two applications of extreme gradient enhancement. From the point of view of the complexity measure, the most accurate method has relatively good levels of resource intensity, as it is 50 times faster than the slowest alternative.

Table 12: Forecast accuracy of the top ten algorithms for incident processing data

Algorithm	**Method**	**Mean Error, ME**	**Root Mean Squared Error, RMSE**	**Mean Absolute Error, MAE**	**Complexity Measure**
Regularized Random Forest	RRFglobal	-23.95	388.82	183.81	2.0%
Regularized Random Forest	RRF	-24.22	389.64	184.59	13.2%
Random Forest	rf	-24.30	389.75	184.61	1.1%
Parallel Random Forest	parRF	-28.04	393.42	188.02	0.7%
Random Forest by Randomization	extraTrees	-23.79	405.33	187.67	5.6%
Random Forest	ranger	-24.27	406.17	188.66	0.3%
eXtreme Gradient Boosting	xgbLinear	-17.56	407.67	188.13	0.3%
eXtreme Gradient Boosting	xgbTree	-10.43	411.34	191.64	0.2%
Boosted Tree	bstTree	-8.33	412.05	185.00	0.1%
Bayesian Additive Regression Trees	bartMachine	-7.16	414.76	196.90	55.5%

If computational optimization is required, it is possible to choose a method with comparably high accuracy and lower computational needs. For the needs of this analysis, we consider the presentation of the most accurate algorithm as satisfactory and we can use it for the needs of risk assessment and management. Ultimately, it is appropriate for organizations to focus only on high-risk transactions within their operational processes. The automated derivation of high-risk observations (transactions or agents) should necessarily result from the application of more than one criterion, with the various criteria and methods ultimately automatically reaching a relative consensus on the anomaly of a case.

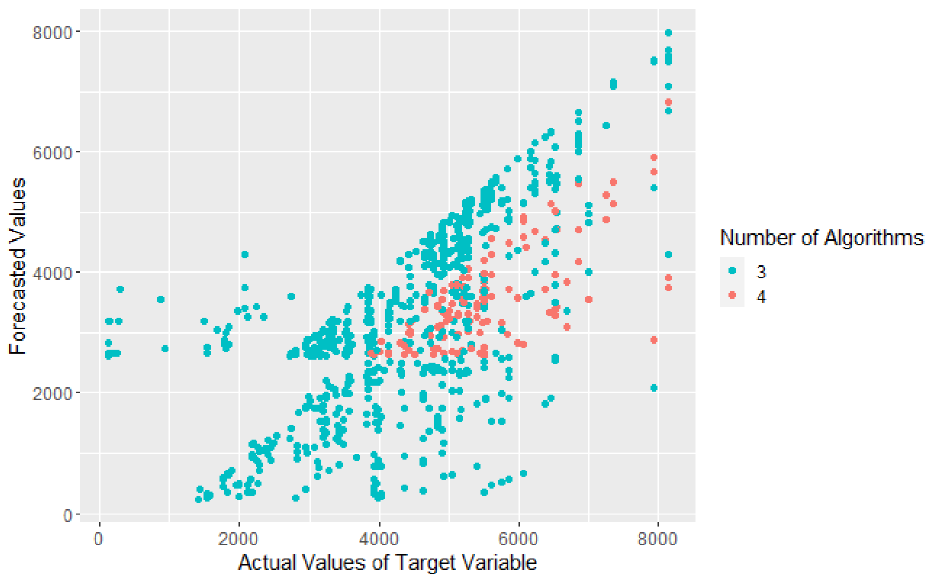

Figure 19: High Risk Anomalous Observations According to Algorithms Consensus

The parallel with a risk management process led by human agents would be the agreement of a group of experts or managers. In this sense, we propose that the criterion for the presence of a potential risk be more than half of the automatic algorithms to classify it as such. Figure 19 graphically shows the risk observations (defined by at least three criteria) as well as the high-risk observations (defined by all four criteria). These are the observations that are subject to active management after their automated identification.

This chapter provides an overview of five main situations in which potential operational risks may arise - in the absence of employees from the workplace, in the valuation of real estate, in the management of communication with stakeholders, in changes in market demand, and when processing signals from customers. The main task of managing these risks is to optimize the balance between potentially negative and potentially positive consequences, and for this purpose we forecast the expected outcomes and look for significant deviations from them. As a potential approach to automation, we propose here to initially apply the general algorithm for managing operational risks and to define the significant deviations among them on the basis of the forecast values. To this end, we propose to use a consensus approach, in which any observation identified as abnormal by more than half of the criteria used is marked as a potential anomaly. The research presented here allows us to outline some main conclusions in this type of tasks.

First, the importance of choosing the appropriate regression algorithm in operational risk management is even clearer. This is clearly visible both in the results presented above and in the ranking distribution of the forecast accuracy. All considered algorithms are ranked in such a way that the one with the highest accuracy, measured by the root of the root mean square error, receives a rank of 1. The algorithm with the second highest accuracy - from 2 and so on to the most inaccurate algorithm, which receives rank of 109. Figure 8 presents a histogram of the averaged ranks of the considered algorithms. There are a small number of algorithms that sustainably generate predictions with a high level of accuracy to all considered tasks. On the other hand, this is not the case with most algorithms, most of which have an average rank between 25 and 75, but we also observe a small number of particularly unsuccessful algorithms with an average rank above 80. This emphasizes the usefulness of applying the statistical algorithm selection stage. before using them in specific scientific or applied problems.

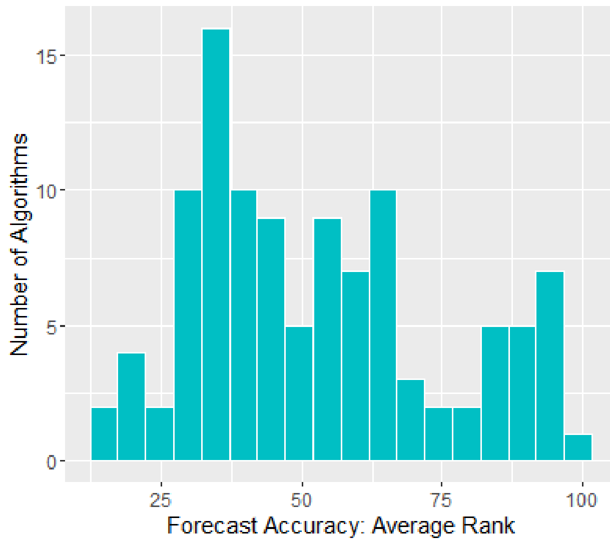

Figure 20: Histogram of Ranked Forecast Accuracy Across All Datasets

Secondly, we note that there are groups of regression algorithms that lead to consistently good results. These are often the different applications of the random forest method, and the estimated values of the different alternatives to random forests are very close to each other. This gives grounds to use highly-optimized variants of the method (eg the ranger implementation in the R language), as they allow significant savings of computational resources in practice without loss of predictive accuracy. Classical statistical methods such as linear regression in general register a much lower forecast value compared to the methods in the field of machine self-learning and it is appropriate to replace them or at least supplement them with more modern approaches. Exceptions in this sense are the situations in which extremely small samples are analyzed. We do not observe a difference between the linear regression model and the alternative algorithms in the field of machine self-learning. On the other hand, the precision measures for such a sample are unstable and less reliable and this result should be interpreted with caution. In any case, the proposed methods have better results when fed with a larger amount of data.

Third, we note that there is virtually no relationship between predictive accuracy and the computational resources required for an algorithm. Figure 9 graphically presents the relationship between the average rank of the algorithm and the logarithm of the time required for its calculations. Visually, no connection is observed between these two characteristics of the considered algorithms. We formally test the relationship by estimating the regression of the mean rank relative to the time required for calculation. Although the time coefficient reaches statistical significance at levels of 5%, its size is extremely small: $\beta = 0.0006$, and the explained variance is only 4%, which shows that there is no high practical significance. This shows that it is appropriate in the automation of the operational risk management process as an additional secondary criterion at the stage of selection of an optimal forecasting algorithm to add its efficiency in terms of necessary computational resources. The almost imperceptible relationship between computational time and predictive accuracy again underscores the ability to choose an algorithm with high accuracy and relatively low computational needs.

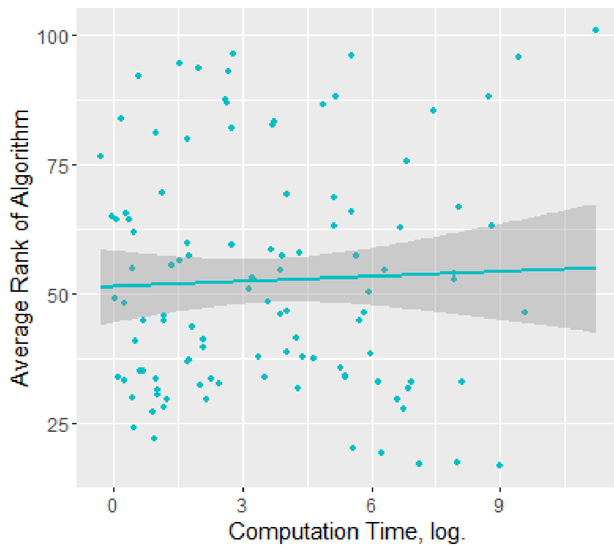

Figure 21: Relationship between Rank of Forecast Accuracy and Algorithm Computation Time (log.)

Fourth, operational risk situations in which the target variable is continuous allow sets of anomaly criteria to be tested. The results presented here show that the proposed approach to consensus between four main criteria is appropriate and workable. We considered the simultaneous application of two statistical criteria (Tucky's formula and the criterion for 4 standard deviations difference), as well as two clustering algorithms based on the density of a given environment (DBSCAN and LOF algorithms). The analysis showed that the farther from the main clusters in the data an observation is given, the more likely it is to be identified as abnormal by more than one criterion. On the other hand, the application of each of the four criteria leads to a very different number of potentially unusual observations, with some of the criteria generating an unrealistically large number of anomalous markings. In order to avoid this problem and minimize the number of false positive classifications, we propose to classify as risky only observations that are such according to at least 3 criteria (or over 50% of the criteria used), and as high-risk - according to at least 4. This approach reduces the number of risk observations and allows for focused management actions.

Chapter 5: Automated System for Operational Risk Management

The automation of the process of operational risk management in the conditions of digital transformation inevitably goes through the construction of information systems that automatically perform the tasks of collecting, processing and analyzing information, as well as supporting decision-making and taking specific management actions. to address potential risks. The quality of the automated analytical processes with which the management information system for assessment and management of the operational risk is charged is a function of the volume and quality of the data with which it is fed. In this sense, the issue of information support of the system is key to the actual implementation of automated risk management. The level of quality of the considered data is imposed as the first most important criterion for using a given data set for the needs of the general algorithm for automated control. Regardless of the data source, they should be highly compliant with quality requirements from both a technological and organizational and business point of view.

Given the high information intensity of the proposed algorithm for automated control, the most important dimensions of quality in this case are the following:

- Accuracy (no errors);
- Objectivity;
- Applicability;
- Appropriate quantity of data;
- Value added.

Only data sets that have satisfactory results against these dimensions are subject to inclusion in the operational risk management MIS. The main sources of this data are two - internal to the organization databases of structured and unstructured data, as well as publicly available data.

The natural completion of the process of automation of operational risks is its successful integration and implementation within a single information system. Based on the general IT architecture framework of the TOGAF Open Group enterprise (see Weismann, 2011; Harrison, 2018), as well as the models proposed by Klein et al. (2016) and other authors (Helu et al., 2017; Hashem et al., 2016; Clement et al., 2017; Theorin et al., 2017; Mayer et al., 2019; Raghupathi & Raghupathi (2014)) we can extend these reference architectures in such a way that they can serve the operational risk management process based on the common automated management algorithm. For this purpose, special modules should be added to ensure the pre-selection of the optimal model, to expand the analytical modules so as to use a real-time optimal model, but also to be able to perform more complex tasks such as sensitivity analysis and simulations. , as well as to expand the scope of the security and system management modules. In addition, the system should include a module to allow risk management once it has been identified. This module supports information about the set of actions to be taken at different risk levels and allows integration into other systems to perform the desired actions (eg termination of a transaction). Schematically, the extended system is shown in Figure 22.

The main contribution of this chapter is the reference architecture of a specialized management information system proposed here. The architecture is designed so that the individual subsystems correspond to the steps of the general algorithm proposed in Chapter 2. The storage and processing subsystem is charged with the functions of information support (steps 2 and 3 of the algorithm). The modeling subsystem is responsible for selecting and evaluating the parameters of the optimal risk management model in each of the situations or tasks to be solved. Its purpose is to automate the training activities of the operational risk management model (steps 4 and 5 of the algorithm). The analysis subsystem automates the model application process (step 6), and the integration with other systems allows for the automation of the control process (step 7 of the algorithm). The continuous improvement process (step 8) is automated through a programmed periodic re-evaluation of the optimal models and their parameters. In addition to these subsystems, the architecture includes three horizontal modules, which are necessary for the smooth operation of the MIS.

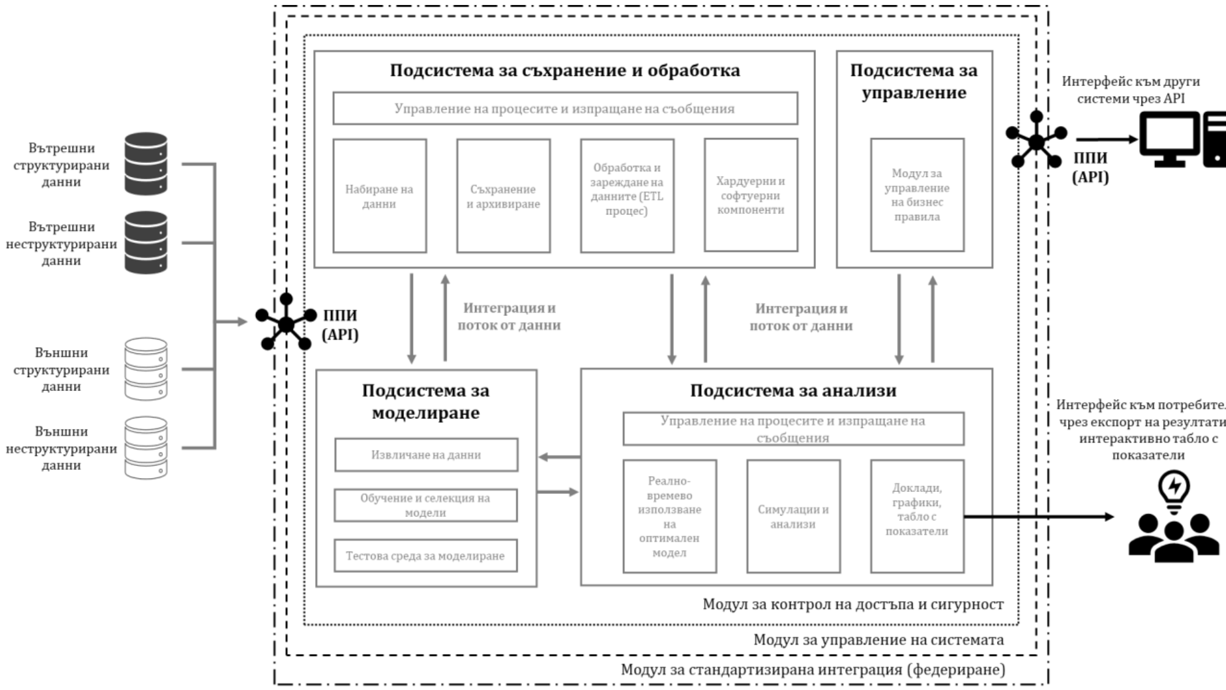

Figure 22: Reference Architecture for a Management Information System for Automated Operational Risk Management

In practice, the vast majority of transactional data in modern information systems are not tagged data. They contain detailed information about the transaction (eg perpetrator and session identifiers, time stamp, nature and details of the transaction itself, execution system, etc.), but individual transactions are marked as risky or not. This is largely due to the fact that most information systems at this stage record transactions, but rarely classify them automatically (Beynon-Davies, 2016; Dumas et al., 2018, pp. 257-277). Therefore, the input data in the automation algorithm is often unmarked data, in which unsupervised training should be performed to search for anomalies. On the one hand, this approach is less accurate than supervised learning (Chandola et al., 2009), but on the other hand it allows an even higher level of automation and integration within an appropriate information system. This is because in the case of unsupervised training, the activities performed by human experts to determine the target variable and its desired values are eliminated and the risk assessment process can proceed completely without human intervention.

We can perform automatic identification and evaluation of potentially anomalous (risk) observations using the already described set of algorithms. First, we use a statistical algorithm for the distance of Mahalanobis. His main idea is to calculate the distance between an observation and the others in multidimensional space and on this basis to identify anomalies (De Maesschalck et al., 2000). It is a well-known fact that these Mahalanobis distances approximately follow the χ2-distribution (Hardin

& Rocke, 2005). Therefore, we calculate the exact levels of significance of the Mahalanobis distance of each observation, comparing it with the theoretical $\chi2$-distribution. We report those observations that have a statistically significant difference from the reference $\chi2$-distribution as anomalous. As a criterion for anomaly, we assume exact significance levels lower than 5%. The second group of approaches that we use to search for anomalous observations within the process of unsupervised self-learning are clustering algorithms based on the density of a given environment. Since we are focusing on contextual anomalies here, it is appropriate to use multidimensional versions of these approaches. In particular, we use the multidimensional version of the DBSCAN algorithm, which aims to identify environments with different densities (Soni & Ganatra, 2016; Hahsler et al., 2019). The basic idea is that anomalous behaviors are found in low-density environments and normal behaviors in high-density environments. The algorithm for local extreme value factor (LOF) works in a similar way, analyzing small environments, on the basis of which it calculates an anomaly measure - LOF (Ma et al, 2013).

As a last potential approach to identify risky behaviors, we can use the analysis of major components (Zenati et al., 2018). It reduces the dimensionality of the data by deriving a certain number of main components. Conventional observations should have high loads on the first major components, while abnormal observations should have higher loads on the lower major components. The distances between observations and components can again be measured using Mahalanobis distances and thus calculate a measure of the abnormality of each individual case (Filzmoser & Todorov, 2013).

Marketing communication through social networks

Given the significant differences between the alternative algorithms, we propose to classify as anomalous observations in which at least three algorithms have a consensus for this. This proposition is based on the main result that the sum of different algorithms generally has better accuracy than any one of them individually (see Makridakis et al., 2020). Figure 23 shows graphically the observations, which are classified as anomalous by three or four algorithms, respectively. The anomalous cases according to three algorithms are 256 in number (3.63% of the total sample), and those that are anomalous according to four algorithms are 92 in number (1.3%). This number is small enough to allow the corresponding more in-depth inspection or in case of impossibility to conduct such - even automatic action by the information system.

We can also look at the profile of the anomalies in relation to the type of status that was posted in the Facebook Live session (see Figure 11). First of all, we note that in hyperlinks (links) we do not observe abnormal behaviors, but only in the other three types of status. They are most pronounced in video materials and text ones (statuses), and the algorithms manage to mark both abnormally high observations - with many shares, and abnormally low ones - with too few. Apart from the links and the photos, we have a relatively common behavior, which shows their sustainability, but also the lack of potential to generate unexpectedly high user engagement. The many anomalies in the video materials reflect their large number, but also significant dispersion. The results obtained so far clearly show that video statuses are characterized by the greatest potential for generating positive risks, but also with the highest level of commitment. In this sense, it is advisable for the organization to further study the video materials with unexpectedly good results and to replicate them, in the process unlocking both business value for themselves and improve the user experience for its customers.

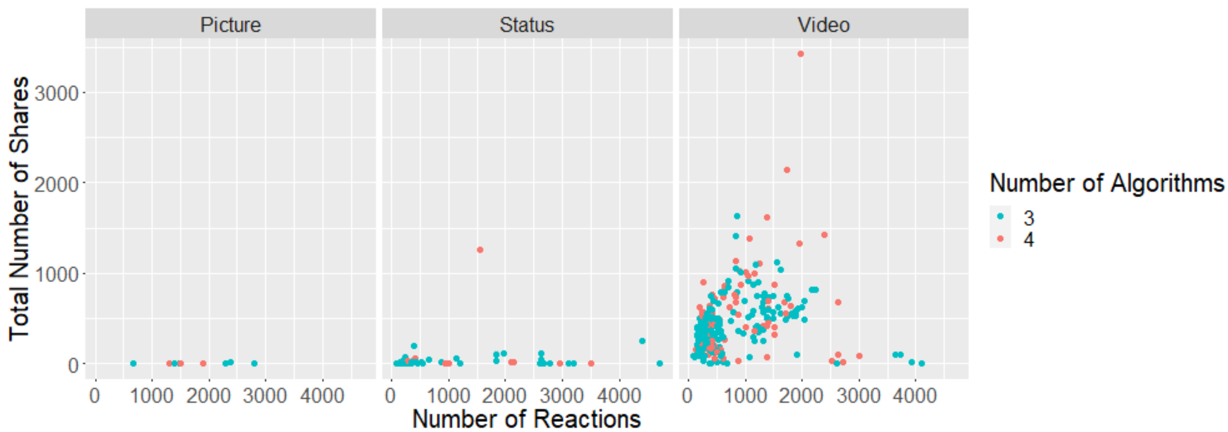

Figure 23: High Risk Anomalous Observations According to Algorithms Consensus

E-shop demand management

The search for potential risks at the individual transaction level carries some challenges. Most remarkably, the problem here is that we aim to identify contextual anomalies, but a significant part of the context is missing. An important part of it is the perpetrator, as certain observations would be abnormal if they were made by one class of agents (eg clients), but completely normal if they were performed by another class of agents (eg employees). In this sense, in risk identification tasks it is always appropriate to work with data for the agents who performed the action, as they can be human agents or automated ones (systems, interfaces, scripts, etc.). In cases where identification of the perpetrator is not possible due to missing data, a transaction-level anomaly approach can be used. In other cases, it is appropriate to use approaches that include both the specifics of the transactions and the specifics of the agents. The analysis of data on trade in goods or services and the identification of extremely high or low values of consumer demand is a classic task in operational risk management. This is important because it allows optimal loading to ensure successful trading, as well as allows the necessary human and organizational resources to be allocated to meet the peaks and to optimize the costs of downturns. In the era of relatively small data sets, this task is solved at an aggregated (summarized) level, considering the total quantities of exchanged goods or services and making decisions about the necessary stocks and employees. In the era of large data sets, low computational costs and flexible statistical algorithms, it is already necessary as a practice to solve the task at the level of individual order or individual user (agent). This allows the risk of extreme behavior to be assessed at a much lower level and at this level to take appropriate action to address it.

As a first step in the analysis of anomalies at the level of an individual agent, it is appropriate to consider the general context of the observed actions. In this case, we have detailed time series of transactions and we can report to what extent their behavior at the aggregate level is normal and whether we observe risks at this higher level. For this purpose, it is appropriate to use methods to search for anomalies within time series. Recall that a time series can be broken down into its main components - long-term trend, seasonal components and random residues (see Cleveland et al, 1990). To apply this approach and choose the optimal model, we direct the reader to Gerunov (2016). We can identify hazardous observations using an appropriate algorithm to identify unusual deviations in the order of random residues in decomposition.

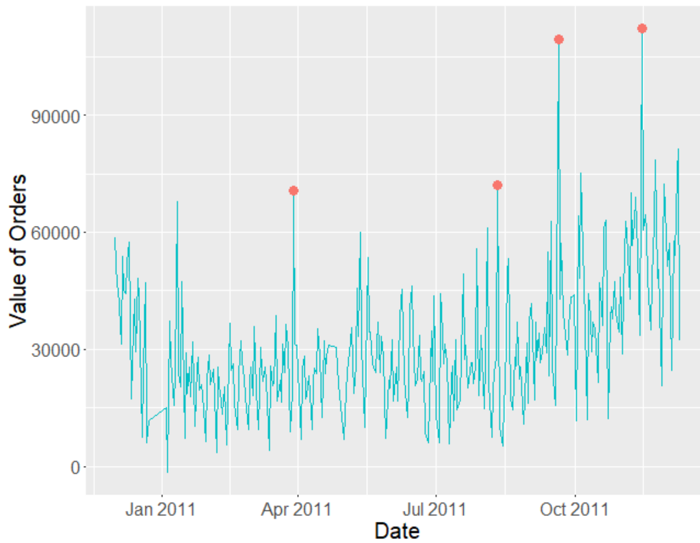

Figure 24: Anomalous Peaks in Aggregated e-Shop Transactions

In Figure 24 we can see the anomalous observations relative to the total daily turnover. The automated algorithm reports four of these (or 1.32% of the sample), all of which are unexpectedly high turnover values. In this sense, these are risks with positive consequences. Although the organization has the ability to take measures to increase their beneficial effect, it is not necessary to take measures to prevent or minimize them. Insofar as there would be a problem, it is related to the possibilities for delivery of the placed orders.

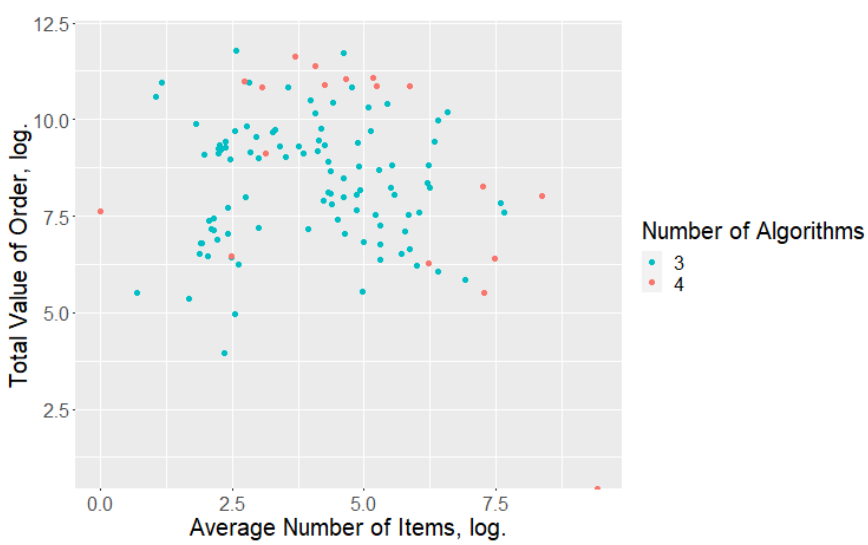

Figure 25: High Risk Anomalous Observations According to Algorithms Consensus

Given the coincidence between these anomalies and abnormal observations at the transaction level, the organization may consider both approaches appropriate for management needs. Applying the consensus approach for agreement between the different algorithms, we mark as potentially anomalous the observations noted by at least three algorithms. These observations, defined as high-risk, are graphically presented in Figure 25. The anomalies identified as such by at least three algorithms are 84 in number or 1.92% of the sample, and those marked by four algorithms - 20 in number or 0.46% of the sample. These proportions show that the consensus method is also successful in this task, successfully identifying a sufficiently small number of observations as high-risk, which are subsequently subject to operational risk management activities.

An additional advantage of this analysis is that these observations are specific individuals that can be further explored or appropriately targeted so that the negative risks to which the organization is exposed are minimized and the positive ones are maximized. Noting that the proposed information system has some characteristics of limited artificial intelligence, it is appropriate to measure the degree of compliance of the proposed system with basic ethical criteria for its useful and safe operation. The summarized parameters of compliance are presented in Table 13, where we assess the degree of compliance with the given criteria on a three-point quality scale (low-medium-high). The assessment is expert and is justified in the last column of the table.

Table 3: Conformance mapping between Bostrom & Yudkowski (2011) ethical criteria and automated MIS for operational risk management

№	Criteria	Conformance Level	Comments on Conformance
1	*Accountability*	Medium	The proposed system is designed so that its design is to reduce the negative effects (a component of net risk) and increase the positive ones (positive deviations from expectations). Although this is done in an organizational context, improving the efficiency of certain organizations also leads to an improvement in overall productivity and the business environment, and hence to an increase in public welfare.
2	*Transparency*	High	The proposed system generates a large amount of visual information and reports that can be used for interim, final and follow-up by human experts. Final decisions are made in a transparent way (consensus between algorithms) and are subject to both intuitive explanation and visualization.
3	*Traceability*	High	The horizontal system management module contains rich functionalities for recording each action and event in the form of log records, as well as for their analysis. The availability of such capabilities allows full traceability of the system's actions both in real time and in the form of ex-post audit.
4	*Sustainability*	Medium	The horizontal module for information security and access management ensures the ability to carefully allocate access levels according to individual agents, user groups and user functions. Additional security features also contribute to higher system resilience, especially to

			external attacks. In order to achieve the highest level here, it is appropriate to implement the system in the information security architecture of the organization.
5	*Predictability*	High	The level of predictability is high, as the system uses sets of known algorithms with some action that can generate reliable solutions. In cases where nondeterministic algorithms are used (eg neural networks or random forests), the design is such that it involves training the models on a large number of samples to ensure the elimination of random stochastic errors. From the point of view of the anomaly criteria, we ensure predictability by maintaining the stability of the models and their parameters.
6	*Avoidance of Harm to Innocent People*	Medium	The proposed automated MIS does not imply integration with systems that have the functionality to inflict physical injuries on human beings. Operational risk management is primarily concerned with the allocation and optimization of resources or processes. Potential damages can occur in case of an incorrect decision of the system, which deprives an individual of resources that he would otherwise have to receive (loan, priority in order, payment, etc.), but does not receive. This can lead to potential tangible and intangible damage, but the goal of the process of continuous improvement of the system (step 8 of the algorithm) is to gradually reduce and minimize these erroneous decisions.

With respect to the six considered criteria of Bostrom & Yudkowski (2011), we identify three with a medium level of compliance and three with a high level of compliance of the system. Given that the proposed automated management information system for operational risk management is not a critical system that manages processes that provide human life or functionalities that threaten it, this level of compliance is sufficient to put the system into productive mode.

IV. Scientific and Applied Contributions

The research conducted in this dissertation creates contributions in three main groups: scientific, scientific-applied and methodological, as the main areas in which they are positioning are economics and science of risk management, econometrics and management information systems.
Among the **scientific contributions** are the following:

- A general algorithm for automated management of operational risks through analysis and synthesis of the existing literature and author's extensions by the method of design science is derived;
- Based on testing of 136 classification algorithms in the field of machine self-learning for those with the highest forecast accuracy in solving operational risk management tasks with a discrete choice;
- Based on testing of 109 regression algorithms from the field of machine self-learning for deriving those with the highest forecast accuracy in solving problems for managing operational risks with a long-term target variable;

- A reference architecture of a management information system is proposed, which fully automates the activities of the derived general algorithm for information risk management and can integrate the tested statistical algorithms.

The **applied contributions** are the following:

- The concept of operational risk has been operationalized in such a way that it is directly applicable to databases;
- A consensus criterion has been proposed for identifying an observation as potentially risky, which relies on the agreement of four different individual criteria in the field of statistics and machine learning;
- Ethical criteria developed for analysis of systems or digital agents with autonomous decision-making are attached to the proposed management information system.

The main **methodological contributions** of the dissertation are as follows:

- 136 classification algorithms and 109 regression algorithms were evaluated and tested, showing how they can be applied to risk management tasks. Most of these algorithms are not applied to this type of task.
- Elements and approaches from the field of design science are used to derive both the generalized management algorithm and the reference architecture of the information system, which shows how design science can be applied to economic tasks of an interdisciplinary nature.

The combination of these contributions leads both to new results in the field of operational risk management and to the creation of a specific digital artifact (architecture of a management information system) that can be used in practice.

V. Relevant Publications

The presented dissertation is tested through publications in the country and abroad, presenting reports at specialized scientific conferences, as well as informal discussions with a number of colleagues working in the field of risk management, economic decision making, economic and business modeling. The publications on the topic of the dissertation are expressed in one monographic textbook, one chapter of a collective monograph, three studies and five articles, as follows:

College Textbook:

1. Gerunov, A. (2017). *Notes on Risk Management*. Sofia University “St. Kliment Ohridski”, Faculty of Economics and Business Administration. ISBN: 978-954-9399-45-5.

Chapter in a Book:

2. Gerunov, A. (2020). Financial and business aspects of the investment in data protection. Chapter 7 in Gerunov, et al. (Eds.), *Privacy by Design: Principles, practices, and technologies*. Sofia University “St. Kliment Ohridski”, Faculty of Economics and Business Administration. ISBN: 978-954-9399-59-2.

Studies:

3. Gerunov, A. (2020). Classification algorithms for modeling economic choice. *Economic Thought*, 2, 45-67.
4. Gerunov, A. (2020). Binary Classification Problems in Economics and 136 Different Ways to Solve Them. *Bulgarian Economic Papers,* 2/2020, 1-31.
5. Gerunov, A. (2019). Risk management: typologies, principles and approaches. *Entrepreneurship*, 7(2), 205-244.

Articles:

6. Gerunov, A. (2020). Machine Learning Algorithms for Forecasting Asset Prices: An Application to the Housing Market. *Economics and Management*, 1, 27-42.
7. Gerunov, A. (2020). Quantitavie approaches to operational risk management in the financial sector. *Annual of the Faculty of Economics and Business Administration, Sofia Univeristy "St. Kliment Ohridski".* (*in print*)
8. Gerunov, A. (2020). Analysis and evaluation of operational risk. *Economic and Social Alternatives*, 2, 24-42.
9. Gerunov, A. (2019). Modelling economic choice under radical uncertainty: machine learning approaches. *International Journal of Business Intelligence and Data Mining*, 14(1-2), 238-253.
10. Gerunov, A. (2016). Automating Analytics: Forecasting Time Series in Economics and Business. *Journal of Economics and Political Economy*, 3(2), 340-349.

The presented publications meet the requirements of Art. 12 of the Development of Academic Staff in the Republic of Bulgaria Act (DASRBA), art. 35 of the Regulations for application of DASRBA (RADASRBA) and fulfill the national quantitative requirements under Art. 1a, para. 1 of RADASRBA for Area 3: Social, economic and legal sciences, Professional field 3.8 Economics.

VI. Conclusion

The results presented in this dissertation have the potential to deepen the knowledge of operational risk management processes, but also to transform them in a way that meets the requirements of the digital transformation. The results presented here, in addition to the available literature and guidelines for future work, have the ambition to help the digitalization of modern organizations in risk management activities in both scientific and applied terms. The aim of this work is to use new approaches, methods and technologies to enhance productivity growth, improve the cost structure and unlock a number of new business opportunities that have an effect on the overall business environment.