REVIEW

By: **Prof. Dr. Stefan Hristov Petranov,** Full Professor at the Faculty of Economics and Business Administration of Sofia University "St. Kliment Ohridski"

Regarding: Review of a dissertation in fulfillment of the requirements for scientific degree **"Doctor of Science"** in Professional Field 3.8 Economics; Economics and Management (Industry) in Sofia University "St. Kliment Ohridski"

Dissertation Author: ***Anton Antonov Gerunov***

Dissertation Topic: ***Automated Approaches to Operational Risk Management***

Legal Basis for the review: participation in the scientific jury for the defense of the dissertation according to Order RD 38-232 / 24.06.2020 of the Rector of Sofia University "St. Kliment Ohridski"

1. Information about the candidate

Anton Gerunov is an associate professor of risk management and quantitative methods for analysis of public policies at Sofia University "St. Kliment Ohridski", where he lectures on data analysis, digital business strategies and management in a digital environment. He has the scientific degree "Doctor of Economics" from Sofia University "St. Kliment Ohridski" Master's Degrees in Economics from Sofia University "St. Kliment Ohridski" in economics and from the University of Stockholm (Sweden) in computer science. His bachelor's degree is from Jacobs University (Bremen, Germany).

Apart from his academic activity, the author of the proposed dissertation also has extensive experience in the public administration as Head of the Office of the Deputy Prime Minister of the Republic of Bulgaria with main field of responsibility: e-government, administrative reform, strategic coordination of OP "Good Governance", and has also served as a consultant and operational director of a company in the technology sector. This combination of academic foundation and expertise from the administration and the technology business creates an excellent basis for the research under review.

2. General overview and relevance of the dissertation

The dissertation consists of 364 pages, of which 332 pages are main text, bibliography and three appendices. The main text is divided into an introductory part, five chapters, a conclusion and three appendices, presenting the classification and regression models tested in the study and the distribution of the samples for these models. The text is illustrated with the help of 141 graphs and 79 tables. The used literature is cited in the bibliographic reference, which includes 410 sources from Bulgarian and foreign authors.

The relevance of the topic of the dissertation is not in doubt. The reason for this is the fact that the modern economic environment is characterized by international integration and globalization, high market dynamics, as well as digitalization and automation, which create conditions for the manifestation of a number of risk factors. At the same time, advances in information technology and the development of analytical methods now make it possible to process large data sets, analyze them and extract information for decision-making with relatively limited resources. In view of these conditions, the author has focused on automated approaches to operational risk management through research on the application of methods in the field of machine learning and their integration into the overall management process.

Some research in this direction are available in both the international and Bulgarian specialized literature, but in the proposed work a significant generalization is achieved by deriving a general algorithm for operational risk management based on author's extensions by the method of design science.

3. Structure

The work has a structure and content that meet the requirements for a dissertation for the scientific degree “Doctor of Science”. It is large enough and contains original works by the author. It is obvious that the author is informed by the practice of the issues under consideration, which is an advantage given their nature. The topic is discussed comprehensively, within a wide scope and with appropriate depth and detail. In general, the work is logically structured, the approach is unified. The author moves from the general theory and typology of risks through to the specification of operational risk and the application of classification and regression models for managing this type of risk. Methods of analysis and synthesis are used in the context of a systematic interdisciplinary approach. The presentation is clear, consistent and a logical connection flows between the individual parts.

4. Dissertation content

Introduction. Before proceeding to the presentation of his research on the merits, in the Introduction the author presents the main characteristics of his dissertation - research thesis, purpose and objectives of the research, scope and methodology, data, and others.

The research thesis is formulated by the author through four components, namely:

1. All activities performed by experts in the operational risk management process may be automated.
2. The application of classification algorithms from the field of machine 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 from the field of machine 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 increased.

Based on the thesis formulated in this way, the aim of the dissertation 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 specialized information system."

The following remarks may be made to the Introduction. On page 7 in the section for research tasks items 5 and 6 the author talks about "the most optimal algorithms". This is illogical because there are no more or less optimal, but simply optimal algorithms. Optimality contains the best, according to a given criterion, and therefore there is no such thing as most optimal. On page 12, when describing the structure of the study (Chapters 3 and 4), the target variable is defined as "continuous" (in Bulgarian: продължителна), and this terminology is subsequently used repeatedly in the text. Probably it is a translation from English, but the term is unfortunate - in the Bulgarian scientific literature the term "continuous" (in Bulgarian: непрекъсната) for this type of variables is ordinary used.

The first chapter logically presents an overview of the literature in the field of risk management. The main typologies of risks in general and

operational risks in particular are considered. The quantitative and qualitative methods for operational risk management are considered, as well as the general strategies and good practices for the process. This chapter argues that there is a need for full or partial automation of the operational risk management process, because although quantitative methods offer high precision and efficiency, in their current version they are excessively dependent on the presence of a relatively large number of qualified experts to be successfully implemented.
I have a note to Chapter One. On page 14, when defining the term risk, an error is made in the formula - instead of the sum of the probabilities participating in it, the expected value of the random variable is incorrectly written.

The second chapter discusses new developments in the field of operational risk management and outlines the main trends in current research. In the context of these developments, the general algorithm for automated management of operational risks is derived. It consists of eight steps, divided 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 algorithmizing of actions so that they can be applied within a specialized information system.
I have no specific notes on this chapter.

Chapter 3 identifies appropriate algorithms for quantifying operational risks in the case of a discrete target variable. For this purpose, 136 different algorithms from the field of statistics and machine learning are tested, and they are applied to solve five classification problems in the field of operational risk.
Among the various tasks, there is a predominantly good performance of the methods from the random forest family, and from the classical econometric methods high accuracy is achieved by the linear discriminant analysis. The results of the study in this chapter give 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.

Several notes can be made regarding this chapter. The text on page 144 needs more clarity. Specifically, it is not clear what a negative and positive class mean and how an algorithm defines a class as positive or negative.
On page 148 it is stated that the complete correlation matrix of the data is presented in the appendices, but there is no such appendix. Also, on page 147, one of the descriptive characteristics of variables is called “skewness”.

Subsequently, this term has been used repeatedly in all tables with descriptive statistics. I guess it is about the relationship between the third central moment of the variable and the standard deviation raised to the third power. If so, it would be better to use the term "asymmetry coefficient", which is usually used in the Bulgarian scientific literature. In Table 32, the variable total liabilities / total assets ratio is presented as a value that averages 0.45. At the same time, by definition in accounting, the total assets of the company are equal to the total liabilities and this ratio should always be 1. So obviously the table is about a different relationship.

Chapter Four follows the logic of Chapter Three but looks at cases where the target task variable is continuous. Both traditional econometric methods and methods in the field of machine 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. The author presents results that show in this case a very good performance of the methods of the random forest family, but also of the neural networks. The chapter also presents results according to which the classical econometric methods of linear regression have a significantly lower performance compared to methods from the field of machine learning and again establishes a relatively weak relationship between forecast accuracy and computational resource intensity.

My notes to this chapter are as follows. There is confusion in the numbering of the sections - from page 208 to page 221 the subsections (written in "italic") start with the number 5, while the main section is with the number 4. In addition, the title of Table 45 and Table 46 is wrong, they do not apply to the data shown in the tables. On page 230 the author concludes that the classical model of linear regression has the weakest results. This statement is exaggerated because it is not confirmed by the data in Table 52. It can be seen that of the five models considered, this model has a 2nd place on the criterion ME, a 3rd place on the criterion RMSE (which in the whole study the author perceived as the principal one), 2nd place according to the MAE criterion and 5th according to the MPE and MAPE criteria, meaning that the results are rather mixed. This conclusion is subsequently summarized on page 266, but a more convincing argument is needed.

On page 217, the author notes that the random forest algorithm that gives the best result leads to the illogical relation that the risk of excessive absence of employees depends statistically significantly on the height of employees. In this case, the comment is limited to one sentence in the sense that this

preponderance of formal criteria over the structural logic of the model is one of the main problems of machine learning algorithms. My opinion is that it would be useful if this problem is considered in more detail. Since the overall thesis of the proposed work is about the advantages and strengths of machine learning algorithms and in this context, especially in a dissertation for a "doctor of sciences", it would be appropriate to comment on their weaknesses, as well. In essence, the purpose of such a model is to serve for risk management, which in the case of the example of employee height cannot have practical value.

The data in Table 55 show descriptive statistics on the cost per unit area for real estate in Taiwan. As a maximum such price (out of 414 observations) is shown 117.5 Taiwan dollars for 1 ping. At the same time, the estimated linear econometric model, according to what is written in the text, shows that increasing the age of the building by one year leads to a reduction in the price by 2740 Taiwan dollars. The increase in the number of nearby stores, in turn, leads to an increase in price by 1208 Taiwan dollars. These results are impossible, with a maximum price of 117.5 dollars, and there is probably some mistake here. In the text of the same page the table is also mistaken - it refers to Table 19, and it should be Table 56.

In Table 61 there is a shift of the rows and the data from the 2nd to the 6th column do not correspond to the variables from the 1st column. If this shift is taken into account, it can be seen that the multiple linear regression model achieves a coefficient of determination 1, regression coefficients 1 for the variables Orders of type A, B, C and 0 for the regression coefficients for all other variables. This is possible if the number of total orders (the predicted variable) is equal to the number of orders of types A, B and C, i.e. if each order in the total number is of one of the three types. From the data shown in the text it is not possible to judge whether this is the case, but if so, it is not correct to predict the total demand through these variables, because the relationship between them and the predicted variable is not regression, but equality by definition. The demand for the orders of type A, B, C should be modeled instead.

Chapter 5 discusses the possibilities for integrating the proposed general algorithm for operational risk management and the corresponding methods for quantitative risk assessment within a specialized information system. A reference information system architecture is derived, which contains four main subsystems - the storage and processing subsystem, the modeling subsystem, the analysis subsystem and the management subsystem. A consensus anomaly

criterion is derived based on four algorithms, with observations classified as anomalous by three of them identified as risky and those classified by four - classified as high-risk. Compliance with ethical requirements for autonomous agents has also been assessed. I have no specific notes on this chapter.

5. Evaluation of scientific and applied contributions

The dissertation has both scientific and applied contributions. Among the former I would like to single out the proposed algorithm for automated management of operational risks and the performed numerous tests with a wide range of models in the field of machine learning. Among the latter, I highlight the proposed consensus criterion for determining risk observations and the ethical criteria developed for the analysis of systems with autonomous decision-making. In general, I accept the contributions listed on pages 332-333.

6. Evaluation of publications on the dissertation topic

The main results of the dissertation are disseminated in the scientific community through publications in authoritative economic publications. A total of ten publications by Anton Gerunov are presented in the materials on the thesis defense procedure. Eight of the publications are in peer-reviewed scientific journals in Bulgaria and abroad. Among the publications are a monograph, three studies, and five articles, one of which is in a journal indexed in Scopus. Five of the publications are in English and the rest - in Bulgarian.

Both quantitatively and qualitatively, the presented publications meet the criteria for a doctoral dissertation, with the following three particularly standing out: the monographic textbook Gerunov, A. (2017). Notes on Risk Management. Sofia University "St. Kliment Ohridski", Faculty of Economics and Business Administration. ISBN: 978-954-9399-45-5; Gerunov Studies, A. (2020). Application of classification algorithms for modeling economic choices. Economic Thought, 2, 45-67, and the article by Gerunov, A. (2019). Modeling economic choice under radical uncertainty: machine learning approaches. International Journal of Business Intelligence and Data Mining, 14 (1-2), 238-253.

The dissertation is original and does not repeat the topic or a significant part of the content of the work for which the educational and scientific degree "Doctor" was obtained. With this the author meets the requirement of art. 76 (2) of the Regulations for the conditions and the order for acquiring scientific degrees and holding academic positions in Sofia University "St. Kliment Ohridski".

7. Evaluation of the summary

The summary presents fully and objectively the contents of the dissertation.

8. Critical comments and recommendations

In Section 4 of this review I have outlined notes regarding the individual chapters of the dissertation. Here I will give a more general comment, as follows: when commenting on the advantages of algorithms based on machine learning, the comparison with multiple linear regression models is not always appropriate, as it is not clear how adequate they are. An illustration in this regard is the model in Table 46. With 740 observations and 21 candidate-explanatory variables, the model achieves an adjusted coefficient of determination of only 0.17. This suggests that either the candidate-explanatory variables are not properly selected, or the relationships between them and the dependent variable are not linear, or both. In this sense, this model is not a convincing basis for comparison.

9. Conclusion

Based on the results obtained and the above arguments, I believe that the proposed dissertation of Anton Antonov Gerunov meets the criteria and requirements set forth in the Rules on the terms and conditions of acquiring scientific degrees and holding academic positions in Sofia University "St. Kliment Ohridski" as well as the Ordinance for Implementation of the Act for the Development of Academic Staff in the Republic of Bulgaria. Thus, I express my **positive opinion** on awarding the scientific degree "Doctor of Science" to Anton Gerunov.

30.8.2020/Sofia

Signed: ……

(Prof. Dr. Stefan Petranov)