

АНОТАЦІЯ

Гуськова В. Г. Методи і методи інтелектуального аналізу даних для оцінювання фінансових ризиків. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки» – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, 2020.

Економічний успіх країн забезпечується застосуванням наукових підходів при управлінні процесами у будь-якій галузі практичної діяльності. Використання сучасних методів математичного моделювання, зокрема, методів інтелектуального аналізу даних до розв'язання проблем прогнозування і моделювання на сьогодні є актуальним завданням. Ризики притаманні будь-якій фінансовій установі або підприємству, а тому належне управління ризиками є важливим аспектом. Керівництво банку або компанії має різний рівень контролю за ризиками, тож внаслідок цього деякими з них можна безпосередньо керувати, а деякі інші ризики знаходяться значною мірою поза контролем управлінців. Найкраще, що можна зробити для запобігання ризиків, – це передбачити їх, оцінити потенційний вплив та бути готовим до реалізації плану реагування на несприятливі події.

Також важливим є оцінювання станів в умовах невизначеностей та за відсутністю методик аналізу розвитку нелінійних нестационарних процесів. Даний підхід дозволяє розв'язати велику кількість складних завдань, які відносяться до різних видів людської діяльності. Перш за все це застосовується у завданнях, які характеризуються великою кількістю чинників або мають при постанові фактори невизначеності. Фінансові ризики відносяться до специфічної сфери кредитно- та ринково-фінансових взаємовідношень банку, фізичних та юридичних осіб, кінцевою метою яких, з точки зору організації, є отримання прибутку. Такі відносини чутливі до

впливу великої кількості різноманітних, часом невизначених ризиків, які у повсякденній роботі фінансової установи мають бути мінімізовані. Основною метою є розробка та вдосконалення математичних методів інтелектуального аналізу даних, які базуються на регресійних моделях, нейронних мережах, мережі Байєса та деревах рішень та відрізняються попередньою обробкою та згладженими початковими даними, що веде до підвищення точності задач фінансових ризиків. Практична цінність роботи полягає у тому, що: розроблено методи і моделі інтелектуального аналізу даних для оцінювання фінансових ризиків. Всі результати роботи доведено до практичного інженерного рівня і впроваджено у навчальний процес інституту прикладного системного аналізу НТУУ «КПІ імені Ігоря Сікорського» та у фінансових організаціях з метою автоматизованого розв'язання задач моделювання, оцінювання і прогнозування втрат у фінансовій компанії як стохастичного багатовимірного процесу. Створено системну методику та систему підтримки прийняття рішень для моделювання, оцінювання і прогнозування ринкового та кредитного фінансових ризиків. Всі теоретичні і практичні результати дисертаційної роботи у повній мірі опубліковано у фахових вітчизняних та закордонних наукових виданнях, що входять до відповідного встановленого переліку, а також виконано їх належну апробацію на міжнародних наукових конференціях і семінарах.

У дисертаційній роботі проаналізовано основні фінансові ризики, управління якими є ключовим фактором, що визначає ефективність фінансової діяльності. Виконано аналіз діяльності банків та інших фінансових організацій, робота яких здійснюється під впливом невизначеностей зовнішнього середовища (ринку, економіки, політики тощо), великої кількості змінних, контрагентів, осіб, поведінка яких не завжди може бути передбачена з прийнятною точністю. Розглянуто можливість мінімізації фінансових ризиків на двох рівнях – на рівні кожної окремої позики та на рівні кредитного портфеля в цілому. В результаті виконання попереднього аналізу встановлено, що найпоширенішими методами оцінювання ризиків для поставленої задачі є

лінійна і логістична регресії, дерева класифікації, нейронні мережі, мережа Байєса. Показано, що для підвищення ефективності прийняття об'єктивних рішень при аналізі кредитного та ринкового ризиків доцільно використовувати мережі Байєса та нечіткі нейронні мережі, які дають можливість враховувати невизначеності ймовірнісного та амплітудного типів. Ці підходи характеризуються швидкими алгоритмами навчання та нескладною інтерпретацією накопичених знань. Такі особливості обраних підходів роблять їх одними з найбільш перспективних і ефективних інструментів моделювання і оцінювання фінансових ризиків.

Розроблено метод оцінювання кредитоспроможності фізичних осіб та запропоновано підхід для вибору підмножини ознак, який характеризується оцінкою значущості результатів за сукупністю підходів та дозволяє підвищити відсоток вірних класифікацій на 3-4% що підтверджується його практичним застосуванням і отриманими результатами експериментальних досліджень. Показано, що процес аналізу кредитоспроможності позичальників кредитів включає в себе розробку методів та критеріїв аналізу процесу кредитування; оцінку потенційного клієнта, а також супровід позики після видачі кредиту. Визначено, що банкам та організаціям для оптимізації кредитного процесу необхідно знаходити компроміс між якістю та ефективністю всього процесу кредитування, а оцінка кредитного ризику здійснюється за допомогою оцінки кредитоспроможності позичальника. Для обчислення нечіткого висновку побудовано нечітку базу знань та використано ННМ в яких результат отримуються на основі нечітких логічних висновків, а параметри функцій належності налаштовуються за допомогою алгоритмів нейронних мереж. Пріоритетною являється задача мінімізації ризику, через що відразу відкидаються ненадійні позичальники, а для всіх інших розв'язується задача максимізації доходу портфеля позик.

Удосконалено метод оцінювання кредитоспроможності позичальника з використанням Байєсівської мережі з урахуванням підвищення якості ймовірнісної моделі та зменшенням величини кредитного ризику.

Розроблено метод оцінювання ринкового ризику на основі інтегрованого застосування ймовірнісної, оптимальної та цифрової фільтрації і регресійної моделі, який відрізняється високою якістю попередньої обробки даних і забезпечує підвищення якості оцінок прогнозів. Запропоновано підхід із використанням попередньої обробки даних у вигляді фільтрації із застосуванням ймовірнісного, оптимального та цифрового фільтрів та регресійної моделі, за результатами якого забезпечено підвищення якості оцінок прогнозів. Виконано експериментальні дослідження із використанням попередньої обробки даних, які показали, що даний підхід в середньому в 2-3 рази ефективніший ніж робота з даними без використання фільтрації. Показано, що процес попередньої обробки даних за допомогою фільтрації є дуже важливим етапом аналізу даних. Застосування методів на цьому етапі, дає можливість значно покращити результати досліджень. Інколи відсутність методів попередньої обробки ставить під загрозу всі подальші кроки по обробці даних. Це може призводити до низької якості результатів, наприклад, оцінки прогнозів характеризуються великими похибками. Проаналізовано та виконано адаптивне прогнозування нелінійних нестационарних процесів, яке є також однією з ключових задач сучасності, у зв'язку з тим, що більшість процесів в економіці, фінансах, екології та технологічних процесах дуже швидко змінюються та не мають єдиного підходу.

Запропоновано підхід із використанням адаптації математичної моделі фінансового процесу до початкових даних ринкових ризиків із застосуванням триконтурної процедури адаптації та побудови моделі нелінійного процесу у вигляді лінійної та нелінійної компонент.

Ефективність усіх розроблених систем керування підтверджена результатами проведених імітаційних моделювань.

Ключові слова: фільтрація, нелінійні процеси, регресійні моделі, лінійні та нелінійні компоненти, триконтурна адаптація, адаптивна Байєсівська мережа, нейронна мережа, прогноз.

ABSTRACT

Huskova V. Data mining methods and models for evaluating financial risks.– Qualifying scientific work on the rights of the manuscript.

Dissertation for obtaining a scientific degree of a Ph.D. degree in specialty 122 « Computer Science». - National Technical University of Ukraine " Igor Sikorsky Kyiv Polytechnic Institute", Ministry of Education and Science of Ukraine, Kyiv, 2020.

Economic growth of countries has provided by scientific approaches to process management in any field of practice. Using of modern methods of mathematical modeling, such as methods of data mining to solve problems with forecasting and modeling today is one of the relevant tasks. The risks are always by any institution or business that's why proper management is an important aspect. The management of a bank or company has different levels of risk control, so as a result, some of them can be directly managed, and some other risks are largely beyond the control of managers. The best thing to prevent risks is to anticipate them, assess the potential impact and be prepared to implement an adverse response plan.

It is also important to assess the states for conditions of uncertainty and in the absence of methods for analyzing the development of nonlinear non-stationary processes. This approach allows you to solve a large number of complex problems related to various human activities. First of all, it is used in tasks that are characterized by a large number of reasons or have uncertainty factors. Financial risks relate to a specific area of credit - and market-financial relations of the bank, individuals and legal entities, the ultimate goal of which, from the point of view of the organization, is to make a profit. Such relationships are sensitive to the impact of a large number of different risks, which should be minimized in the daily work of a financial institution.

The main aim of the thesis is to develop and improve mathematical methods of data mining based on regression models, neural networks, Bayesian networks and

decision trees and characterized by pre-processing and smoothed initial data, which leads to increased accuracy of financial risks.

In the work the following scientific results were obtained: The method of market risk assessment based on the integrated application of probabilistic, optimal and digital filtering and regression model has been developed, which is characterized by high quality data pre-processing and improves the quality of forecast estimates. An adaptation method of the mathematical model of the financial process to the data has been developed, which differs in the applying of the three-loop adaptation procedure and provides construction of the nonlinear process model in the form of linear and nonlinear components. Method of enhancing the quality of forecasting borrowers' solvency has been improved, which differs with the combined approach to the selection of repressors and the use of alternative forecasting methods, which provides optimization of weighting coefficient estimates for individual forecasts. The method of assessing the borrowers' solvency using an adaptive Bayesian network has been improved, which is characterized by increased adequacy of the probabilistic model and provides a reduction in the amount of credit risk.

The practical significance of the thesis results is in the developed methods and models of data analysis for financial risk assessment. The results of the thesis fulfillment were used in educational process of the MMSA Department of Institute for Applied System Analysis (IASA) at the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", and in the financial organizations in order to automatically solve the problems of modeling, estimating and forecasting losses in a financial company as a stochastic multidimensional process. A systematic methodology and decision support system for modeling, assessing, forecasting market and credit financial risks has been created. All theoretical and practical results of the dissertation published in professional domestic and foreign scientific journals included in the relevant list, as well as their proper approbation at international scientific conferences and seminars.

In the work analyzed the main financial risks, the management of which is a key factor that determines the effectiveness of financial activities. Implemented an analysis

of the activities of banks and other financial organizations, whose work is carried out under the influence of environmental uncertainties (market, economy, politics, etc.), a large number of variables, counterparties, individuals whose behavior can not always be predicted with acceptable accuracy. The possibility of minimizing financial risks at two levels has been considered - at the level of each individual loan and at the level of the loan portfolio as a whole.

As a result of preliminary analysis, it was found that the most common methods of risk assessment for the task are linear and logistic regressions, classification trees, neural networks, Bayesian network. It is shown that to increase the efficiency of making objective decisions in the analysis of credit and market risks, it is advisable to use Bayesian networks and fuzzy neural networks, which make it possible to take into account uncertainties of probabilistic and amplitude types. These approaches are characterized by fast learning algorithms and simple interpretation of accumulated knowledge. Such features of the chosen approaches make them one of the most promising and effective tools for modeling and assessing financial risks.

In the dissertation work the comparative analysis of efficiency for using statistical methods, logit and probit models, Bayesian networks, decision trees, the neural network with the indistinct logical conclusion of Mamdani and Sugeno which showed the efficiency of fuzzy neural networks credit analysis for individual borrowers. The estimation method of solvency for individual borrowers has been developed. An approach is proposed for selecting a subset of features, which is characterized by assessing the significance of the results on a set of approaches and allows increase substantially the level of correct classifications by 4-5%, which is confirmed by its practical use and the results of experimental studies.

It is shown that the process of credit analysis of borrowers includes the development of methods and criteria for analyzing the lending process; assessment of a potential client, as well as support of the loan after the credit. It is determined that banks and organizations need to find a compromise between the quality and efficiency of the entire lending process to optimize the credit process, and credit risk is assessed by assessing the borrower's creditworthiness. To calculate the fuzzy inference, a fuzzy

knowledge base is constructed and FNN are used in which the result is obtained on the basis of fuzzy logical inferences and the parameters of membership function are configured using neural network algorithms. The priority task is to minimize the risk, which immediately rejects unreliable borrowers, and for all others, the task of maximizing the income of the loan portfolio.

The improved method was proposed for assessing the solvency of a credit borrower using the Bayesian network, and taking into account improving the quality of the probabilistic model and reducing the amount of credit risk. The method of market risk assessment based on the integrated application of probabilistic, optimal, and digital filtering and regression model has been developed, which is characterized by the high quality of data pre-processing, and improved quality of the forecast estimates.

An approach using pre-processing of data in the form of filtering with the use of probabilistic, optimal, and digital filters and a regression model is proposed, as a result of which the quality of forecast estimates is improved. An approach using pre-processing of data in the form of filtering using probabilistic, optimal and digital filters and a regression model is proposed, as a result of which the quality of forecast estimates is improved.

Experimental studies were performed using pre-processing of data, which showed that this approach is on average 2-3 times more efficient than working with data without the use of filtering. It is shown that the process of pre-processing data by filtering is a very important step in data analysis. The use of methods at this stage makes it possible to significantly improve the results of research. Sometimes the lack of pre-processing methods jeopardizes all further steps in data processing. This can lead to poor quality results, for example, estimates of forecasts are characterized by large errors. Adaptive forecasting of nonlinear non-stationary processes is analyzed and performed, which is also one of the key tasks of today, due to the fact that most processes in economics, finance, ecology and technological processes change very quickly and do not have a single approach.

The efficiency of all the developed control systems is confirmed by the results of the simulations.

Keywords: filtering, forecast, nonlinear processes, regression models, linear and nonlinear components, tricycle adaptation, adaptive Bayesian network, neural network.

Список публікацій здобувача:

***Статті у періодичних наукових виданнях держав, які входять до
Організації економічного співробітництва та розвитку та/або
Європейського Союзу***

1. Zaychenko Yu.: Recognition of Objects on Optical Images in Medical Diagnostics Using Fuzzy Neural Network Neffclass / Yu. Zaychenko, V. Huskova // International Journal "Information Models and Analyses". 2016. – №5. pp. 13-22. *(Болгарія). (запропоновано підхід для розпізнавання предметів на медичних зображеннях у медичній діагностиці).*

***Статті у наукових фахових виданнях України, які входять до
міжнародних наукометричних баз даних***

2. Tymoshchuk O. L. A combined approach to modeling nonstationary heteroscedastic processes / O. L. Tymoshchuk, V. H. Huskova, P. I. Bidyuk. // Radio Electronics, Computer Science, Control. – 2019. – №2. – С. 80–89. *(включено до наукометричної бази Web of Science). (запропоновано модифіковану методологію моделювання нелінійних нестационарних процесів та схему адаптації для побудови моделей).*

Статті у наукових фахових виданнях України

3. Бідюк П. І. Аналіз кредитоспроможності за допомогою методів інтелектуального аналізу даних / П. І. Бідюк, В. Г. Гуськова // Електронне моделювання. - 2019. - Т. 41, № 2. - С. 111-120. *(досліджено підхід до мінімізації ризику платоспроможності позичальника для банківської системи та інших*

фінансових компаній).

4. Гуськова В. Г. Оцінювання кредитоспроможності позичальників кредитів методами інтелектуального аналізу даних / В. Г. Гуськова, П. І. Бідюк. // Міжнародний науково-технічний журнал «Системні дослідження та інформаційні технології. – 2019. – № 2. – С. 31–48. *(проаналізовано основні методи математичного моделювання і оцінювання кредитних ризиків, запропоновано математичні моделі для аналізу кредитних ризиків індивідуальних позичальників на основі альтернативних методів).*

5. Гуськова В. Г. Розробка сценарного підходу на основі моделей інтелектуального аналізу даних / В. Г. Гуськова, П. І. Бідюк // Наукові праці Донецького національного технічного університету. Серія : Інформатика, кібернетика та обчислювальна техніка. – 2016. – № 2. – С. 158-164. *(запропоновано підхід на основі логістичної регресії для аналізу кредитоспроможності позичальників з використанням пакета RStudio із аналізом та застосуванням фактичних вихідних даних для оцінювання кредитних ризиків у майбутньому).*

6. Гуськова В. Г. Система підтримки прийняття рішень для прогнозування фінансових процесів на основі принципів системного аналізу / В. Г. Гуськова, Данилов В.Я., П. І. Бідюк, О.Л. Жиров // Міжнародний науково-технічний журнал «Системні дослідження та інформаційні технології. – 2019. – №1. – С.20–36. *(досліджено концепцію розв'язання задач адаптивного прогнозування на основі методології системного аналізу).*

7. Гуськова В. Г. Аналіз кредитоспроможності позичальників кредитів за допомогою логістичної регресії. / В. Г. Гуськова, П. І. Бідюк // Наукові праці Донецького національного технічного університету. Серія: Інформатика, кібернетика та обчислювальна техніка. – 2017. –№2. - С. 54-60. *(побудовано математичні моделі аналізу для коротко-, середньо- та довгострокових прогнозів; розроблено авторегресійні рівняння).*

Наукові праці, які засвідчують апробацію матеріалів дисертації

8. Бідюк П.І., Гуськова В.Г.: Застосування нечітких правил регресійного аналізу до фінансових даних, Institute for Modelling in Energy Engineering, NAsc of Ukraine, NAsc of Ukraine, September 12-14, 2018, Kyiv, Ukraine.

9. Huskova V.H., Bidyuk P.I.: Estimating financial risk using systemic approach, Проблеми інформатизації, тези доповідей шостої міжнародної науково-технічної конференції, 14 – 16 листопада 2018 року, Черкаси – Баку – Бельсько-Бяла – Харків.

10. Bidyuk P., Huskova V., Terentiev O.: Client solvency estimation using intellectual data analysis approach, The VIth International Conference «Advanced Information Systems and Technologies, AIST 2018», 16-18 May 2018, Sumy, Ukraine.

11. Zaychenko Yu., Huskova V.: Application of fuzzy neural network nefclass for recognition of medical images in diagnostics, System Analysis and Information Technologies; 18-th International Conference SAIT, 30 May - 02 June 2016, Kyiv, Ukraine.

12. Huskova V., Bidyuk P.: A Combined Approach to Modeling Heteroscedastic Processes and Financial Risk Estimation, Всеукраїнська науково-практична конференція комп'ютерна інженерія і кібербезпека: досягнення та інновації 27–29 листопада 2018 року, Кропивницький, Україна.

13. Гуськова В.Г., Бідюк П.І.: Побудова сценаріїв із використанням байєсівських методів, міжнародна науково-технічна конференція "моделювання і комп'ютерна графіка", 18–24 вересня, 2017 року, Покровськ, Україна.