Investments play a significant role in the functioning and development of the economy. Risk management is an integral part of the formation of the investment portfolio. This means that an investor must be willing to take on a certain level of risk in order to receive a certain level of return. However, when forming an investment portfolio, an investor faces such problems as market unpredictability, asset correlation, incorrect asset allocation. Therefore, when forming an investment portfolio, an investor should carefully study all possible risks and try to minimize them. The object of research is an approach to risk management in the formation of an investment portfolio using the method of reinforcement training. The basic principles of formation of the investment portfolio and determination of risks are described. The application of the method of reinforcement training for building a model of risk management of investment portfolio is considered. The process of selecting optimal investment assets based on alternative data sources that minimize risks and maximize profits is also considered. A functional model of the process of risk optimization in the formation of an investment portfolio based on machine learning methods has been developed. The functional model constructed makes it possible to build a process of risk optimization, including asset selection, risk comparison and assessment, to form an investment portfolio and monitor its risks. The study results showed that the proposed approach to the formation of the investment portfolio increased the total growth of the investment portfolio by 0.4363 compared to the base model. Also, the volatility indicator improved compared to the market, as evidenced by the percentage difference between the initial and final cash amount, which increased from 128.98 to 295.57

Keywords: investment portfolio, risk management, machine learning, actor-critic, learning without a trainer

UDC 004.891.2
DOI: 10.15587/1729-4061.2023.277997

DEVELOPING A RISK MANAGEMENT APPROACH BASED ON REINFORCEMENT TRAINING IN THE FORMATION OF AN INVESTMENT PORTFOLIO

Vitalii Martovytskyi
Corresponding author
PhD, Associate Professor*
E-mail: vitalii.martovytskyi@nure.ua
Volodymyr Argunov
Postgraduate Student*
Igor Ruban
Doctor of Technical Sciences, First Vice-Rector**
Yuri Romanenkov
Doctor of Technical Sciences, Professor
Department of Management
National Aerospace University «Kharkiv Aviation Institute»
Chkalova str., 17, Kharkiv, Ukraine, 61070
*Department of Electronic Computers**
**Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

Received date 07.02.2023
Accepted date 18.04.2023
Published date 28.04.2023

How to Cite: *Martovytskyi, V., Argunov, V., Ruban, I., Romanenkov, Y. (2023). Developing a risk management approach based on reinforcement training in the formation of an investment portfolio. Eastern-European Journal of Enterprise Technologies, 2 (3 (122)), 106–116. doi: https://doi.org/10.15587/1729-4061.2023.277997*

1. Introduction

Investments play a significant role in the functioning and development of the economy, and changes in physical volumes and quantitative ratios of investments affect the volume of public production and employment, the development of industries and sectors of the economy.

The active development of the world financial market testifies to the acquisition of special importance of financial instruments in the system of the global economic mechanism for the functioning of the global economy. With the intensification of globalization processes, portfolio investment goes through different stages of evolution and acquires new characteristics [1, 2].

Global foreign direct investment (FDI) flows declined by 35 percent in 2020, reaching USD 1 trillion, from USD 1.5 trillion in 2019 (Fig. 1) [3]. This is the lowest level since 2005 and almost 20 % below the 2009 minimum after the global financial crisis. Lockdowns around the world in response to the COVID-19 pandemic have slowed down existing investment projects, and the prospect of recession has forced multinational corporations (MNCs) to reevaluate new projects. The fall in FDI was much sharper than the fall in gross domestic product (GDP) and trade.

That is why industrial companies, concerns, and holdings need to manage individual innovative projects and portfolios and optimize the risks associated with them.

The portfolio theory of Harry Markowitz, published in 1952, is considered classic in this area. As part of his theory [4], Markowitz first expressed the idea of the need to measure, track, and control not only profitability but also risk. As quantitative metrics describing interesting characteristics of portfolio assets, Markowitz proposed using expected returns and risk levels, which are estimated based on a history of asset price fluctuations. A key aspect of Markowitz theory is not only the idea of diversifying a portfolio in order to reduce the overall level of risk but also the formulation of its quantification. One drawback is that Markowitz's analysis did not involve short positions (negative values of asset weights).

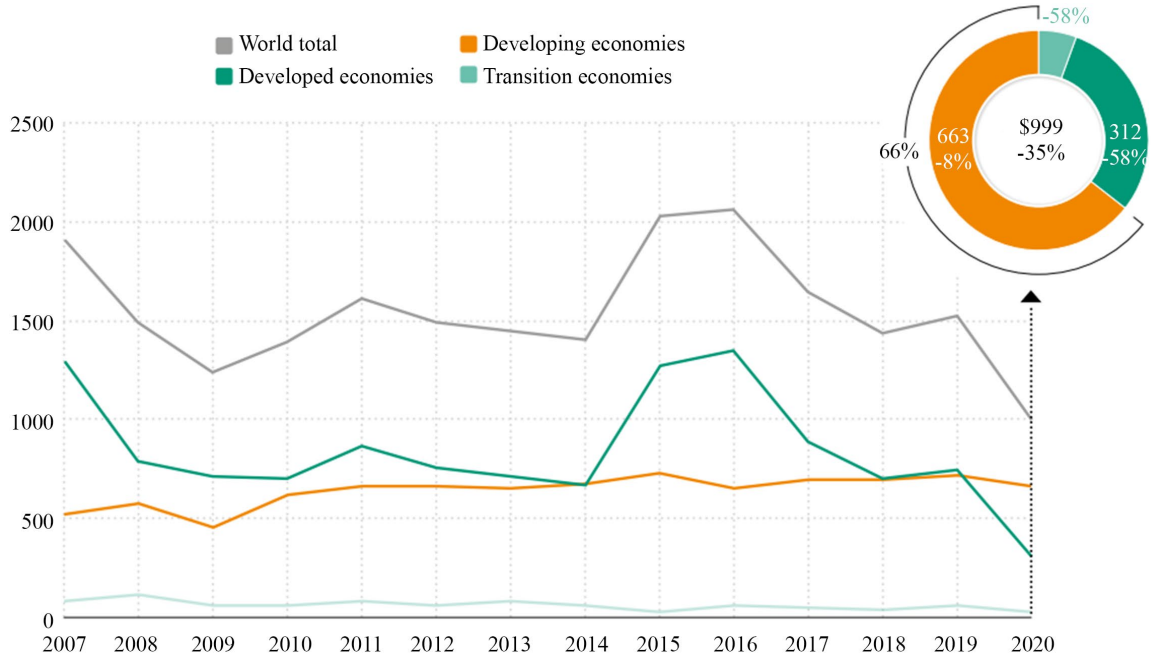

Fig. 1. The flow of foreign direct investment in the world and by groups of economies, 2007–2020 (USD billions and interest rates)

Similar studies were conducted by Roy; a similar approach was published in [5], which is also based on data on expected return and portfolio variance. When forming his model, Roy was guided by the principle «Safety First principle» [5, 6], that is, the expected profitability should not be less than the predefined level. And unlike Markowitz's theory, Roy's model made it possible to take into account short positions in the formation of a portfolio.

A common problem in both approaches is the need to estimate the expected return and variance of assets, as well as their correlation structure. The easiest way to assess these points is to estimate based on historical data. However, in some cases, this approach can lead to significant valuation errors and, as a result, to inefficient indicators of the formed portfolio in the future.

Risk management in the formation of project portfolios is necessary to solve the following management tasks:

– formation of a balanced portfolio of projects, taking into account its compliance with the company's goals and ensuring the necessary balance between risk and return on investment;

– risk management of portfolio projects, including building an effective risk management system in the company;

– ensuring the necessary transparency and attractiveness of the company to investors, insurance companies in order to reduce the cost of attracted financing, reduce the cost of insurance programs, improve credit ratings and the value of the company.

Unsupervised learning methods such as cluster analysis, amplification learning, and deep neural network training can help investors identify basic market behavior patterns and build an optimal portfolio.

For example, cluster analysis can help investors divide assets into clusters according to their interaction and the risks associated with these assets. Investors can review this information and build a portfolio that reflects different clusters, providing a balance between risk and potential return.

Reinforcement training and deep neural network training can help investors identify patterns of market behavior and develop risk management strategies based on these patterns.

Consequently, trainerless learning methods can be useful to investors in practical use when forming an investment portfolio. It is important that the study contains specific recommendations that would be accessible and understandable to investors.

Thus, the development of methods and means of risk management in the formation of investment portfolios is an urgent task that requires continuous improvement.

2. Literature review and problem statement

The formation of an investment portfolio, on the one hand, is aimed at preserving capital at the expense of conditionally risk-free assets, and on the other hand, at increasing it by including risky assets. Unlike monoinvestment, portfolio investment makes it possible to improve investment conditions by giving a set of assets an investment characteristic that are unattainable from the position of a single asset. The main investment characteristic that interests any investor is the ratio of risk and portfolio return. Finding a balance between these indicators, depending on individual investment goals, is the main task of the theory of investment portfolio management [3, 7, 8]. An unambiguous approach to the formation of an optimal portfolio in financial theory does not exist [9].

One of the most well-known methods of risk assessment is value at risk (VaR) [9]. It is a generalizing quantitative statistical measurement of risk, which makes it possible in one number to generalize the influence of different risk factors and takes into account the correlation between the influence of risk factors. VaR characterizes the amount that will not exceed the expected losses during a certain period with the predefined probability. The VaR indicator was first used by JP Morgan to improve risk efficiency. Taking into account the peculiarities of regulation and the impact of various factors on financial markets, the effectiveness of the VaR methodology is perceived ambiguously. Despite the classic nature of use, the parametric method for calculating VaR needs to be improved.

In [10], a practical example of risk management based on rebalancing is given. In addition to risk management methods based on diversification and hedging, the rebalancing method is used to reduce the risks of the investment portfolio. Unfortunately, this approach to risk management and reduction is not always possible. In [11], a new framework State-Augmented RL is presented. Its structure is aimed at solving two unique problems of financial project management:

– heterogeneity of data – the information collected for each asset is usually diverse, noisy, and unbalanced;

– environmental uncertainty – the financial market is multifaceted and unstationary.

To include heterogeneous data and increase resilience to environment uncertainty, SARL supplements asset information with a forecast of their price movements in the form of additional states.

In [12], the authors presented a flexible approach to the formation of an investment portfolio at the industry level and used the Merton model of conditional claims to assess the impact of the carbon tax shock on the market value of equity and debt instruments. In the process, calibrating the model using detailed company-level vulnerability data. As a result, a decrease in the market value of banks' assets by 2–13 % of fixed capital was revealed with a tax shock of EUR 100 and an increase to 6–29 % with a tax shock of EUR 200. But the results of this approach can only be used as an additional factor for building a general model for forming an investment portfolio. In [13], the authors reported a study aimed at introducing the ELECTRE-TRI and FlowSort methods when choosing a stock portfolio as one of the most popular and important subjects for decision-making. They also compare the results of each method to understand how these methods work in the tasks of forming investment portfolios. In this study, the best worst method was used to determine the weights of the criteria. Four approaches for ELECTRE-TRI were considered. In ELECTRE-TRI, if the results of the pessimistic approach differ greatly from the optimistic approach, there are several incompatibilities between categorical portfolio formation. According to the optimistic or pessimistic approach, the nine alternatives belong to the best (first) or worst (third) class but not to the intermediate (second) class. Thus, for the correct formation of the portfolio, additional information may be required, for example, the inclusion of other decision-making criteria or the provision of accurate estimates. The result also shows that using the veto threshold for ELECTRE-TRI does not give a good result for this task.

Based on the analysis, we can conclude that such classical methods as presented in [9] cannot be fully used to optimize the risks of forming an investment portfolio. This is due to the fact that under real conditions, in addition to classical financial indicators and indicators on the attractiveness of certain assets, many external factors play a role.

In place of classical approaches, the latest methods and approaches [10–13] are gradually beginning to be introduced, based on information technology. These methods and approaches try to introduce additional indicators and parameters with the help of which it is possible to assess in a certain way the influence of certain external factors in optimizing the risks of forming an investment portfolio. But such methods and approaches are rather highly specialized and/or aimed at certain types of investments, or at certain factors that affect the risk assessment of certain types of assets.

That is why research and development of risk management methods in the formation of an investment portfolio become relevant.

The transition from post-industrial to information society, which takes place in the XXI century, cannot be imagined without intensive information exchange and development of information systems. The key communicative role in it belongs to networks that freely form an association of people and interest groups. Communication technologies make it possible to create social communities (Internet community) with almost any given characteristics – educational, professional, age. They are formed against the background of acceleration of social time and strengthening the dynamics of communication forms in the process of social reproduction. At the same time, stable relations give way to constant changes, and society becomes similar to reflective and communication communities [14, 15].

Therefore, our study proposes an approach to risk optimization in the formation of an investment portfolio, which will include, in addition to classical indicators, the impact of social media on asset price volatility.

This choice of additional indicators is due to the fact that by analyzing social media, you can get not only information from reputable investors or public figures but various discussions that are related to a particular asset.

The method of reinforcement training is an effective approach for building an investment portfolio. This approach is that an agent (for example, an investor) interacts with the environment (for example, the stock market) and makes decisions based on the rewards received (profit or loss).

In the case of forming an investment portfolio, the agent can choose different assets (for example, stocks, bonds, funds, etc.) and distribute his/her capital among them according to his/her goals and strategy. The agent can receive rewards in the form of investment profits and penalties in case of losses.

The vast majority of reinforcement learning (RL) and neurodynamic programming (NDP) methods fall into one of the following two categories:

– methods intended only for subjects (actors) work with a parameterized policy family. The performance gradient with respect to the actor's parameters is estimated directly by modeling, and the parameters are updated in the direction of improvement [16–20]. A possible disadvantage of such methods is that gradient estimates can have greater variance. Moreover, as the policy changes, the new gradient is evaluated regardless of previous estimates. Consequently, there is no «learning» in the sense of accumulating and consolidating old information;

– methods intended only for critics rely solely on the approximation of the value function and aim to obtain an approximate solution to the Bellman equation, which proposed an almost optimal policy [21, 22]. Such methods are indirect because they try to optimize the policy space directly. The method can succeed in constructing a «good» approximation of the value function but with reliable guarantees in terms of the near optimality of the resulting policy;

– actor-critical methods aim to combine the strengths of only actor and critical methods. The critic uses intermediary architecture and simulation to explore the value function, which is then used to update the actor's policy settings towards performance improvements. Such methods, if based on a gradient, may have desirable convergence properties, unlike critical methods, for which convergence is guaranteed under very limited conditions. They promise to provide faster convergence (by reducing variance) compared to methods based only on actors. On the other hand, the theoretical understanding of actor-criticism methods was limited to the

possibility of presenting policies in search tables [23]. That is why the work proposes such an approach to learning.

3. The aim and objectives of the study

The aim of this work is to develop an approach to risk management in the formation of an investment portfolio based on machine learning methods. This will provide an opportunity to analyze the impact of social media on asset price volatility. In turn, this will make it possible to comprehensively assess the risks in the formation of the investment portfolio and on the basis of this build a system of support and decision-making in the formation of the company's investment portfolio.

To accomplish the aim, the following tasks have been set:

– to build a functional model of the process of risk optimization in the formation of an investment portfolio based on machine learning methods;

– to conduct an experimental study on the proposed approach.

4. The study materials and methods

The object of research is the process of risk management in the formation of an investment portfolio based on reinforcement training.

The main hypothesis of the study assumes that the use of reinforcement learning methods will effectively manage risks in the formation of an investment portfolio, reducing costs and increasing portfolio profitability. The developed approach to risk management will reduce costs in portfolio formation and increase portfolio profitability.

The algorithms of actors' criticism as algorithms of stochastic gradient in the actor's parameter space were considered. When the actor's parameter vector is θ, the critic's job is to compute the projection approximation Π_θ. The actor uses this approximation to update his/her policy in the approximate direction of the gradient.

The actor-critic is similar to the policy gradient algorithm called REINFORCE [24] with a basic level. Reinforcement is Monte Carlo learning, indicating that total income is taken from the full trajectory. But in actor's criticism we use bootstrap. So, the main changes will be in the action-value function, and it will take the form:

$$A(s_t, a_t) = \sum_{t'=0}^{T-1} r_{t'} - b(s_t). \quad (1)$$

$b(s_t)$ was replaced by a function of the value of the current state. Then it can be represented as follows:

$$A_{\pi_\theta}(s_t, a_t) = r(s_t, a_t) + V_{\pi_\theta}(s_{t+1}) - V_{\pi_\theta}(s_t). \quad (2)$$

The state value $V_\pi(s)$ is the expected total reward starting at state s and operates in accordance with policy π.

If the agent uses the predefined policy to select actions, the corresponding value function is defined as:

$$V_\pi(s) = E\left[\sum_{i=1}^{T} \gamma^{i-1} r_i\right] \forall s \in S. \quad (3)$$

The optimal state function has a high possible value function, compared to another value function for all states:

$$V^*(s) = \max_\pi V_\pi(s) \forall s \in S. \quad (4)$$

In RL, if we know the optimal value function, then the policy corresponding to the optimal value function is the optimal π^* policy.

Alternatively, the preference function is the error of TD, as shown in the Actor-Critic scheme in Fig. 2.

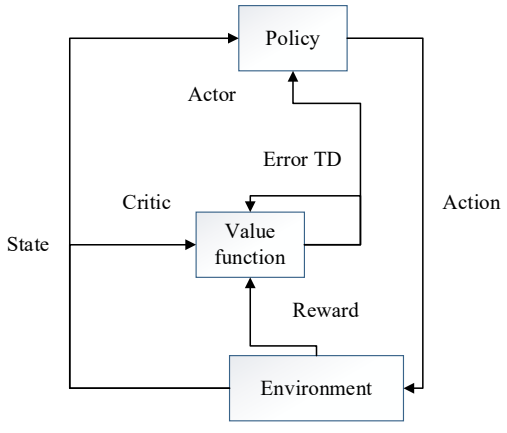

Fig. 2. Algorithm actor-critic scheme

The expression of the actor's policy gradient can be expressed as follows:

$$\nabla J(\theta) \approx \sum_{t=0}^{T-1} \nabla_\theta \log \pi_\theta(a_t, s_t) A_{\pi_\theta}(a_t, s_t). \quad (5)$$

The actor-critic algorithm [25] can be represented as follows:

Step 1. Take an example (s_t, a_t) that uses a π_θ policy from a network of actors.

Step 2. Evaluate the preference function A_t. This can be called an error of TD. In the actor-critic algorithm, the preference function is created by the critics network using formula (2).

Step 3. Estimate the gradient by expression (5).

Step 4. Update the θ policy settings using the formula:

$$\theta = \theta + \alpha \nabla J(\theta). \quad (6)$$

Step 5. Update the weights based on RL criticism (Q-learning). δ_t corresponds to the preference function:

$$w = w + \alpha \delta_t. \quad (7)$$

Step 6. Repeat steps 1 to 5 until we find the optimal policy π_θ.

Taking into account the peculiarities of reinforcement learning methods, the following simplifications were adopted in the current study:

– lack of consideration of macroeconomic factors in the formation of the portfolio;

– lack of consideration of individual needs and limitations of investors when forming a portfolio.

5. Results of investigating the risk management approach in the formation of the investment portfolio

5. 1. Functional model of risk management process in the formation of an investment portfolio

Investment risks are understood as the reasons for the volatility of investment returns. All investments are subject to

different risks. The greater the volatility of prices, the higher the level of risk. Understanding the risks involved in owning different securities is crucial to building the right investment portfolio. Probably, it is the high risk that discourages many investors from investing in stocks and forces them to keep money in so-called risk-free savings deposits, certificates of deposit, and bonds.

After analyzing the methods of risk assessment in the formation of the investment portfolio [3–14], it can be concluded that approaches to risk assessment in the formation of the investment portfolio should include three categories of indicators:

1) classic financial data that can be obtained from public reports of companies;

2) technical financial data (indicators), which make it possible to obtain data on future prices using quotes data for a certain period of time;

3) alternative non-financial data, which include the following indicators:

– economic and technological threats: inflation, economic sanctions, rising prices for resources;

– political and legal threats, for example, imperfection of legislation in a certain region where the company's assets are concentrated;

– socio-demographic threats, such as reduced purchasing power;

– potential threats from social media: analysis of public opinion, analysis of messages of key figures-leaders in social networks.

Taking into account all these indicators, a general structure of the approach to risk optimization in the formation of the investment portfolio was formed.

To describe the approach to risk management in the formation of an investment portfolio, it is proposed to use the following tuple:

$$RM = \{S, Actor, Critic, RL\}, \quad (8)$$

where S is the set of environmental states, formed on the basis of data collected in the process of monitoring asset states; *Actor* – an asset management agent for the formation of an investment portfolio, designed to assess the risk of the i-th asset and make decisions to include an asset in the portfolio; *Critic* – an agent whose task is to approximate the Q-function – the utility function of control; *RL* is a reinforcement learning algorithm that forms an optimal risk management policy when forming an investment portfolio. During training, the system (*Actor, Critic*) learns by interacting with some environment.

Each element of the state of the medium is described by the following tuple:

$$S = \{fin, sem\}, \quad (9)$$

$$fin = \{opA, opC, am, lq, pr, st, dbt, mt\}, \quad (10)$$

where *sem* i-th set of indicators as a result of semantic analysis of social media resources.

A *fin* is a set of classical financial indicators, where:

opA is the set of operational analysis indicators;

opC – set of indicators of operating costs;

am – a set of asset management indicators;

lq – multiple liquidity indicators;

pr – a set of profitability indicators;

st – a set of capital structure indicators;

dbt – multiple indicators of debt service;

mt – set of market indicators.

In turn, *Actor* and *Critic* are two independent neural networks, where $\pi(s,a,\theta)$ is a policy function that controls the actions of our agent and $\hat{q}(s,a,w)$ is a value function that measures how good these actions are.

Since we have two models (*Actor* and *Critic*) that need to be trained, this means that we have two sets of weights (θ for our action and w for our critic) that must be optimized separately:

$$\Delta\theta = \alpha\nabla_\theta\left(\log\pi_\theta(s,a)\right)\hat{q}_w(s,a). \quad (11)$$

$$\Delta w = \beta\left(R(s,a) + \gamma\hat{q}_w(s_{t+1},a_{t+1}) - \hat{q}_w(s,a)\right)\nabla_w\hat{q}_w(s,a). \quad (12)$$

At each step t we take the current state (S_t) from the environment and pass it as input through our Actor and Critic. Politicy adopts a state, decides on an asset (A_t) and receives a new state (S_{t+1}) and a reward (R_{t+1}), Fig. 3.

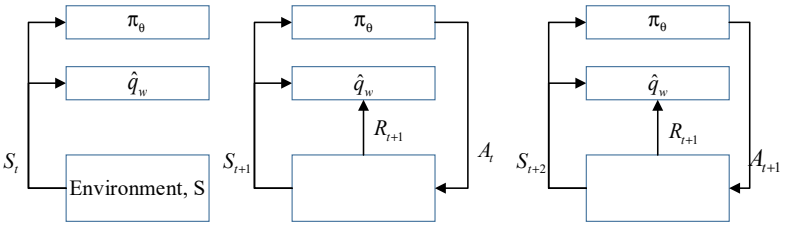

Fig. 3. Diagram of the process of teaching actor's criticism

Due to this:

– the critic calculates the value of performing this action in this state;

– the actor updates his/her policy (weight) parameters using the q value.

Thanks to its updated parameters, the Actor performs the next action to be done in A_{t+1} given the new state S_{t+1}. The critic then updates his/her parameters.

Since value-based methods have high variability, it is proposed to replace the value function with a modified one (2).

This function will report an improvement compared to the average value of the action taken in this state.

In other words, this function calculates the additional reward we get if we perform this action. An additional reward is something that goes beyond the expected value of that condition.

If $A(s,a) > 0$: the gradient shifts in that direction.

If $A(s,a) < 0$ (our action is worse than the mean of this state), the gradient shifts in the opposite direction.

5. 2. Experimental research on the proposed approach

The following set was used to conduct the experiment:

$$\left\{\begin{array}{l} date,\ asset,\ open,\ high,\ low,\ close,\ volume, \\ djcp,\ turbulence,\ MACD\_12\_26\_9, \\ MACDh\_12\_26\_9,\ MACDs\_12\_26\_9, \\ ADX\_14,\ DMP\_14,\ DMN\_14, \\ CCI\_14\_0.015,\ RSI\_14,\ mentions, \\ mentions, tweets\_count, tweets\_normalized, \\ sentiments\_negative,\ sentiments\_neutral, \\ sentiments\_positive \end{array}\right\}, \quad (13)$$

where *date* – date (period); *asset* – ticker (exchange code) of the asset; *open* ∈ *fin* – starting price of the asset for the period; *high* ∈ *fin* – maximum asset price for the period; *low* ∈ *fin* – minimum asset price for the period; *close* ∈ *fin* – the final price of the asset for the period; *volume* ∈ *fin* – the total amount of the asset participating in trading for the period; *adjcp* ∈ *fin* – adjusted closing price of an asset reflecting the value of that asset after taking into account any corporate actions such as payment of dividends, divestitures, mergers, etc.; *turbulence* – index of financial turbulence; MACD_12_26_9 – MACD indicator/line MACD; MACDh_12_26_9 – MACD indicator/bar chart; MACDs_12_26_9 – MACD indicator/signal line; ADX_14 – ADX indicator/trendline; DMP_14 – ADX indicator/positive directional movement; DMN_14 – ADX indicator/negative directional movement; CCI_14_0.015 – CCI indicator; RSI_14 – RSI indicator; *mentions* – the total number of mentions in social text messages that were collected and analyzed for tonality over a certain period; *tweets_count* – the same *mentions*, only «tweets», messages on the social network Twitter, without retweets, reposts, and other «social activity», only original messages; *tweets_normalized* – daily average from polarity, where polarity is a superposition of *negative*, *neutral*, *positive* tonality analysis estimates in the range {–1, ..., 1}; *sentiments_negative* ∈ *sem* – aggregated negative score based on the results of the analysis of sentiment of social text messages related to the asset for a certain period; *sentiments_neutral* ∈ *sem* – aggregated neutral score based on the analysis of the tone of social text messages related to the asset for a certain period; *sentiments_positive* ∈ *sem* is an aggregated positive score based on the results of analyzing the tone of social text messages related to an asset for a certain period.

The parameters when setting up the model were selected as follows:

– *lr*=0.0001 – classic learning rate;

– *vf_loss_coeff*=0.5 – loss coefficient of function «value»;

– *entropy_coeff*=0.01 – coefficient of regularizer entropy (control of freedom);

– *model_fcnet_hiddens*: [512, 512] – size of agent cascade.

When implementing the model, the Ray framework was used, both for building an agent model and for distributed tuning/training, plus an upgraded framework for creating an environment for OpenAI Gym, taking into account the financial features of the model. Plus, a mini framework was written to compile datasets and calculate technical indicators using configuration files to speed up this process. Backtesting was also implemented with Quantopian's Empyrical in mind.

All experiments were conducted on 2–3 virtual machines (Google Cloud VM) with 8 cores/16 GB of RAM. For tuning, we used: 500 iterations for each grid search element, 1.5–2 hours per element, for training: 1000 iterations, 3.5–4 hours.

To test the performance of the proposed approach and conduct the experiment, several models were trained, while the architecture of the model, environment, etc., is the same, the difference is in the composition of the dataset:

1) full dataset without alternative financial data;

2) a complete dataset with alternative data from Twitter;

3) full dataset with alternative data from the news.

In this case:

– random seed was configured in all possible modules of the model where it is possible, and it does not disrupt the learning process of the model for maximum similarity at all stages and reproducibility of the whole process;

– a sufficiently large number of iterations were chosen so that the possible differences were completely insignificant since the possible ways of exploration process came to the same optimal solutions thanks to other fixed SEED-s.

Tuning of hyperparameters of the model took place on a full dataset without alternative financial data to achieve good results of the model as a whole. The experiment was conducted using the same hyperparameters of architecture and environment in order to be able to prove or disprove the real effect of alternative non-financial data on improving the quality of the model.

An important point of training took place on the so-called rollout fragments, these are randomly selected fragments from the entire time series. Thanks to this approach, both «bias» and memorization of the best choice are excluded, as well as retraining, which is although minimally possible in such systems. But for the sake of purity and reliability, this kind of comparative experiment should be excluded.

After training the models, we carried out:

– classical validation, in the form of loss control, vf_loss, and other standard metrics of the learning process. But in this situation and for the purposes of the experiment they are completely useless, therefore they were not included in the results of the experiment, they were only tested to confirm the correct learning process;

– for the main and additional experiment, a basic analysis of the model performance and several basic backtesting metrics were conducted;

– for the main experiment, full backtesting of models and analysis of relevant results were carried out.

Next, paired graphs will be shown describing how the comparison was made and the key points about them for all experiments. Plus, all backtesting metrics and its specific graphs for the main one. The order is the first without alternative data, the second with them.

Since older «alt-data» data does not exist or is not possible to obtain in principle but, given the results, it is for the better. The training was conducted on data obtained from sources starting in 2022 and early 2023, that is, during a period of severe global recession in markets, so some stability indicators like MDD, Downside Deviation, Stability, Beta have low and similar indicators.

Global backtest metrics (in basic model/alt-data format) are given in Tables 1, 2.

Table 1

Backtesting results — global metrics for the base model

Metrics	Value	Metrics	Value
Gain	128.98	Omega	1.51
CAGR	1.34	Volatility	0.42
Sharpe	2.27	Return	1.34
Tail	1.07	Stability	0.84
Sortino	3.48	Downside	0.27
MDD	–0.23	Beta	1.18
Calmar	5.87	Alpha	1.68

Table 2

Backtesting results — global metrics for alternative data

Metrics	Value	Metrics	Value
Gain	295.57	Omega	1.96
CAGR	3.11	Volatility	0.43
Sharpe	3.5	Return	3.11
Tail	1.16	Stability	0.92
Sortino	5.93	Downside	0.26
MDD	–0.23	Beta	1.13
Calmar	–0.18	Alpha	3.73

Metrics for comparison with the market (Alpha&Beta) are calculated compared to S&P-500.

Fig. 4, 5 show a plot of the effectiveness of the base model and the proposed approach with alternative data.

Fig. 6, 7 show a plot for calculating the Sharpe sliding coefficient.

The Sharpe coefficient of a strategy is designed to measure the average excess profitability of a strategy as the ratio of volatility «sustained» to achieve this return. This is a broad measure of the ratio of reward to risk of strategy.

The Sharpe moving coefficient on an annualized basis simply calculates this value based on trading data for the previous year. It provides a constantly updated, albeit retrospective, view of the current reward-risk ratio.

A low Sharpe coefficient (below 1.0) means that significant yield volatility is maintained at a minimum average return.

The negative Sharpe coefficient implies that it would be better to have an instrument representing the risk-free rate used in the calculations. In this case, not only is the average return of the strategy lower than that achieved by the risk-free rate, but the volatility of these inefficient returns also remains. This is a good indicator of the effectiveness of a strategy.

Fig. 8, 9 show a plot for calculating the sliding coefficient Alpha.

From the plots, you can see a significant increase in peak values, some redistribution of intervals, some intervals have a smaller peak but longer duration.

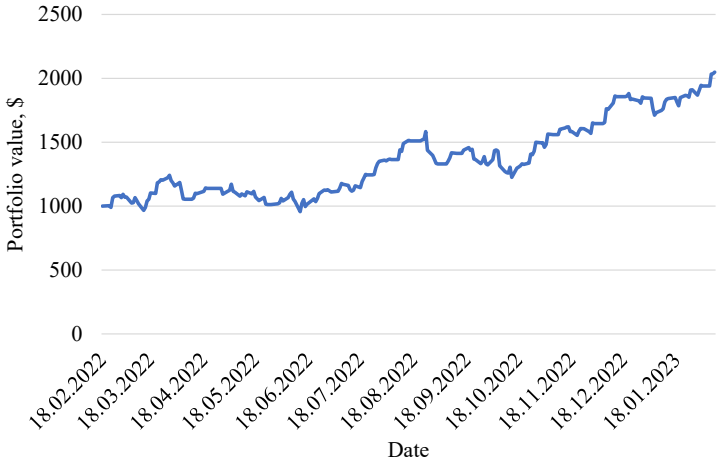

Fig. 4. Baseline model performance plot

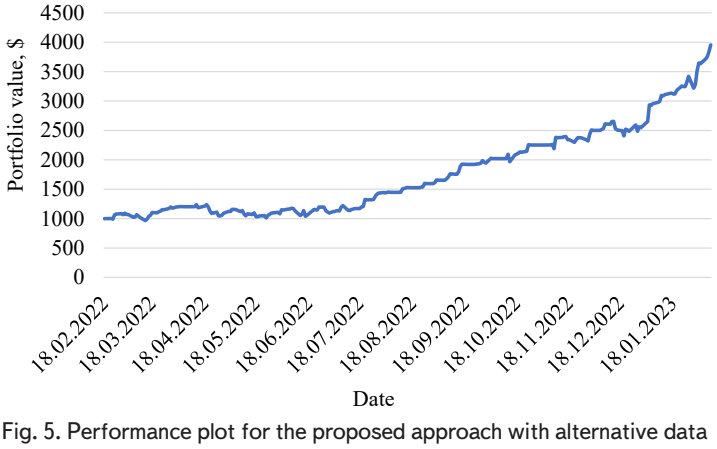

Fig. 5. Performance plot for the proposed approach with alternative data

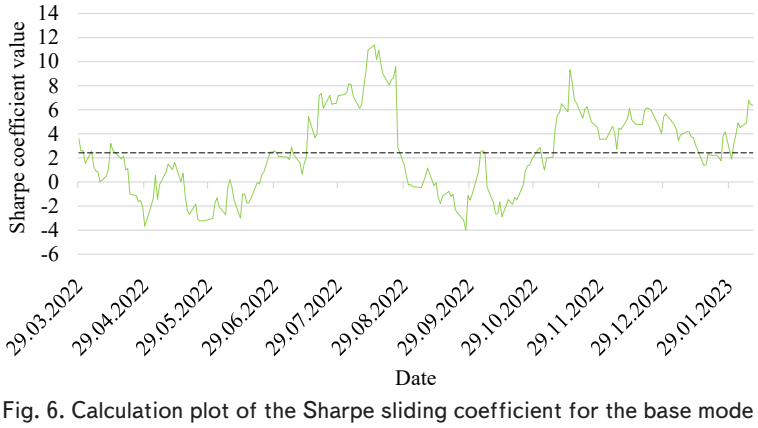

Fig. 6. Calculation plot of the Sharpe sliding coefficient for the base mode

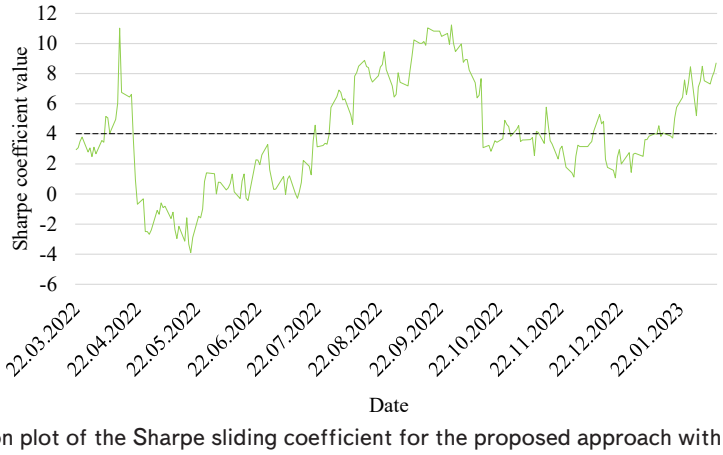

Fig. 7. Calculation plot of the Sharpe sliding coefficient for the proposed approach with alternative data

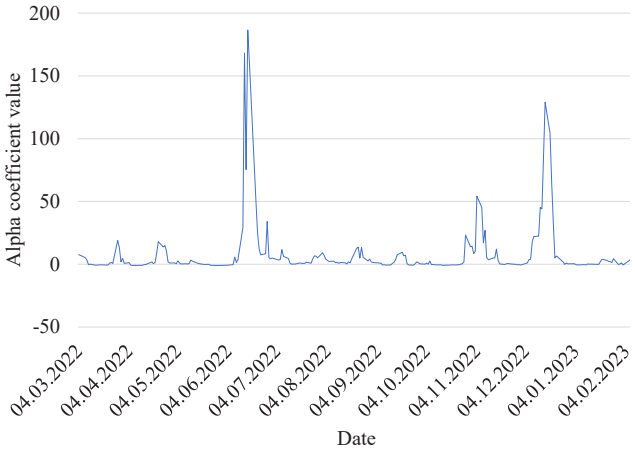

Fig. 8. Alpha sliding coefficient calculation plot for base model

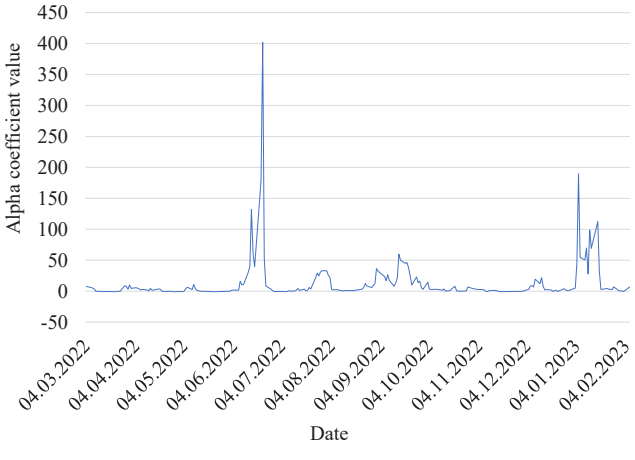

Fig. 9. Calculation plot of sliding coefficient Alpha for the proposed approach with alternative data

The plots shown in Fig. 10, 11 present a comparison of the moving factor Beta.

Despite the high similar peaks, it can be seen that the number and steepness of these peaks is generally lower.

Fig. 12, 13 show the sliding window calculation plots for the Rolling Drawdown coefficient.

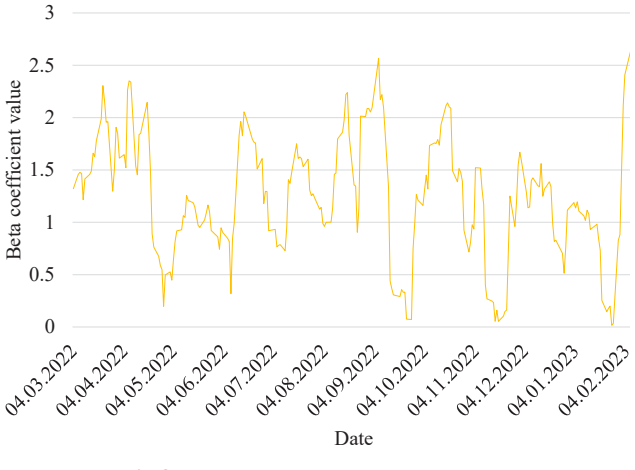

Fig. 10. Sliding coefficient Beta calculation plot for the base model

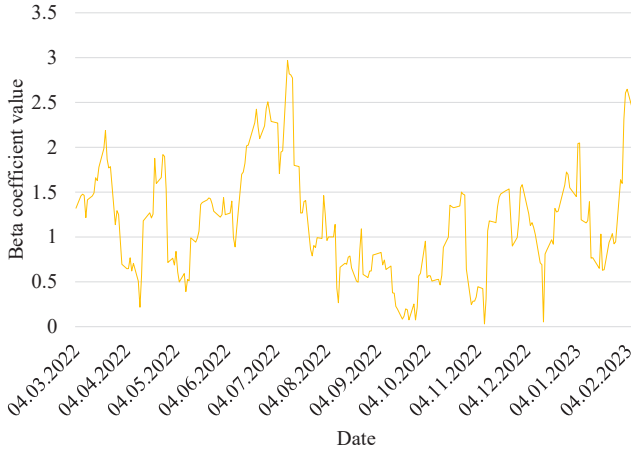

Fig. 11. Calculation plot of the sliding coefficient Beta for the proposed approach with alternative data

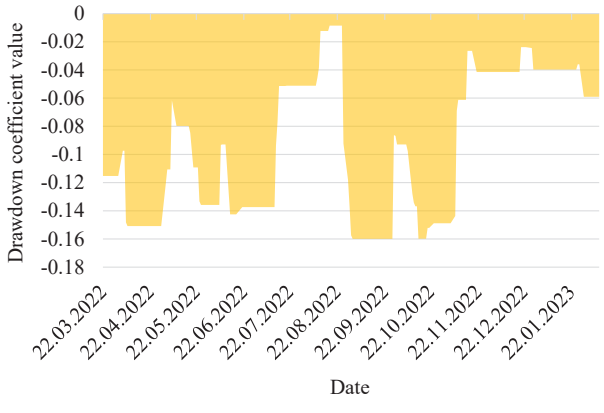

Fig. 12. Rolling Drawdown sliding factor calculation plot for the base model

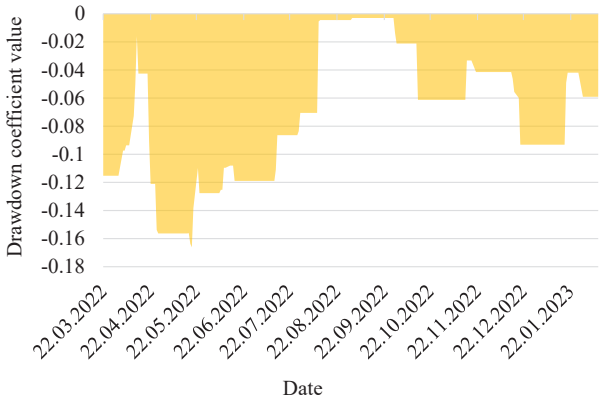

Fig. 13. Rolling Drawdown sliding coefficient calculation plot for the proposed approach with alternative data

Despite the high similar global indicators, it is seen that the number of intervals in which they have high values has decreased.

Fig. 14, 15 present plots for calculating the sliding window for the Cumulative Returns coefficient.

On the charts you can see a noticeable decrease in losses and, as a result, improved stability and profitability. There is also an increase in growth against the market.

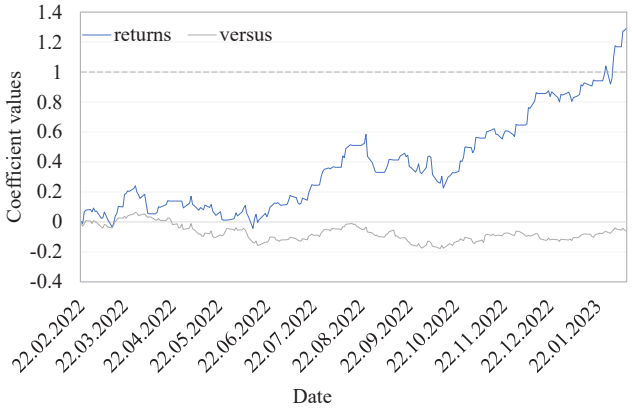

Fig. 14. Cumulative Returns sliding factor calculation plot for the base model

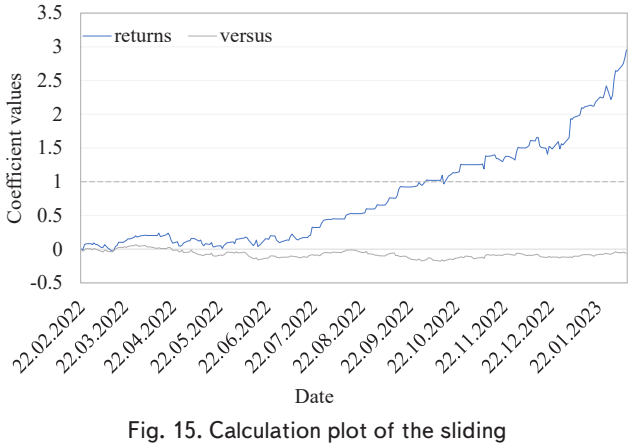

Fig. 15. Calculation plot of the sliding coefficient Cumulative Returns for the proposed approach with alternative data

6. Discussion of results of investigating the approach to risk management in the formation of the investment portfolio

After analyzing the general metrics of the backtest, which are presented in Tables 1, 2, and on the plots in Fig. 4–15, the following improvements can be seen in the formation of the investment portfolio:

– having considered the Gain indicator, you can see that the total growth of the investment portfolio increased by 0.43637717;

– having considered the CAGR indicator, you can also see that the average annual increase also increased by about 0.43;

– having considered the indicator Sharpe Fig. 6.7 it can be noted that the index as a whole is very high since with such fluctuations the losses are not so great, but the application of the proposed approach significantly improves it;

– having considered the Sortino indicator, you can also notice an improvement from the coefficient value for the base model 3.48 to 5.93;

– after analyzing the MDD coefficient, you can see that the level of subsidence improved by 0.05. While not a significant improvement, this is a significant improvement for this metric;

– having considered the indicator that measures the return of the portfolio adjusted for risk (Calmar coefficient) Tables 1, 2 you can see an increase in efficiency in relation to risk.

On the plots of profit growth (portfolio value history) Fig. 4, 5 one can see a marked reduction in losses and, as a result, stability and profitability improved.

From Fig. 6, 7, one can see an increase in the number of intervals with higher values and a decrease in the number of intervals with negative values and, as a result, an improvement in the average and global values of the metric.

After analyzing the plots in Fig. 8, 9, it is noticeable that some intervals have a smaller peak, which can generally be considered an improvement since in this case, the stability and indicator are higher for a longer time.

The plots shown in Fig. 10, 11 also demonstrate high similar peaks and the number and steepness of these peaks is generally lower than that of the base model, indicating an improved indicator and less instability.

According to the plot of the history of aggregate profit growth Fig. 14, 15 there is a marked reduction in losses and, as a result, the increased stability and profitability. You can also see an increase compared to the market.

The results of the study showed that RL agents can significantly improve asset allocation because they outperform strong baselines in contrast to the methods demonstrated in [11–13].

Although the experiment successfully proved the work of the proposed approach, there are cases when the result does not correspond to the expected. This is caused by shadow factors, such as insider transactions, speculation in the market, and other unfair actions of interested parties, which also significantly affects the state of assets.

The proposed approach is of great practical importance for investors and asset managers. Risk management is an important aspect of investing because it helps reduce possible losses and ensure sustainable portfolio returns. The use of reinforcement learning methods makes it possible to identify and analyze the risks that arise when forming a portfolio and develop optimal strategies for their management.

In addition, the developed risk management approach reduces portfolio management costs as it provides more accurate and efficient asset allocation solutions. This is important because management costs can affect overall portfolio returns.

Consequently, the risk management approach based on reinforcement learning has great practical potential for investors and asset managers, allowing them to reduce risk and increase the return on their portfolio.

As a shortcoming, it can be noted that a small number of content sources were used during the experiment and the result of the analysis may not actually fully reflect the real picture. From the side of the approach itself, the problem is solved quite simply by adding additional sources of content but on the technical side, it will require additional labor and computing resources.

For further development of this approach, attention should also be paid to the formation of validated and extended data sets that can be used in training neural networks.

7. Conclusions

1. A functional model of the process of risk optimization in the formation of an investment portfolio based on machine learning methods has been built. The developed functional model makes it possible to build a process of risk optimization, including asset selection, risk comparison and

assessment, building an investment portfolio and monitoring its risks. In addition, various constraints can be taken into account, such as asset limits, risk limits, and portfolio value limits. The use of machine learning methods in the development of a risk optimization model makes it possible to use a large amount of data and take into account various factors that affect the risks and returns of the investment portfolio.

2. Our experimental study on the proposed approach to the formation of the investment portfolio showed that the total growth of the investment portfolio increased by 0.43637717 compared to the base model. Also, the volatility indicator improved compared to the market, as evidenced by the percentage difference between the initial and final amount of cash, which increased from 128.98 to 295.57.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

The data will be provided upon reasonable request.

References

1. Kaftia, M. A. (2019). The Formation of Modern Portfolio Theories: the Main Problems and Tendencies of Development. Business Inform, 2 (493), 414–419. doi: https://doi.org/10.32983/2222-4459-2019-2-414-419
2. Romanenkov, Y., Vartanian, V. (2016). Formation of prognostic software support for strategic decision-making in an organization. Eastern-European Journal of Enterprise Technologies, 2 (9 (80)), 25–34. doi: https://doi.org/10.15587/1729-4061.2016.66306
3. Zadoia, A. O. (2019). Portfolio investments in Ukraine: chance or challenges? Academic Review, 2, 81–92. doi: https://doi.org/10.32342/2074-5354-2019-2-51-8
4. Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7 (1), 77–91. doi: https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
5. Roy, A. D. (1952). Safety First and the Holding of Assets. Econometrica, 20 (3), 431. doi: https://doi.org/10.2307/1907413
6. Sullivan, E. J. (2011). A.D. Roy: The Forgotten Father of Portfolio Theory. Research in the History of Economic Thought and Methodology, 73–82. doi: https://doi.org/10.1108/s0743-4154(2011)000029a008
7. Jagannathan, R., Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. The Journal of Finance, 58 (4), 1651–1683. doi: https://doi.org/10.1111/1540-6261.00580
8. Ratushna, Yu. S. (2019). Foreign financial investment development factors. Naukovyi visnyk Uzhhorodskoho natsionalnoho universytetu. Seriya: Mizhnarodni ekonomichni vidnosyny ta svitove hospodarstvo, 24 (3), 59–66. Available at: http://nbuv.gov.ua/UJRN/Nvuumevcg_2019_24%283%29__13
9. Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. Economic Review, Federal Reserve Bank of San Francisco.
10. Sunchalin, A. M. et al. (2019). Methods of risk management in portfolio theory. Available at: https://www.revistaespacios.com/a19v40n16/a19v40n16p25.pdf
11. Ye, Y., Pei, H., Wang, B., Chen, P.-Y., Zhu, Y., Xiao, J., Li, B. (2020). Reinforcement-Learning Based Portfolio Management with Augmented Asset Movement Prediction States. Proceedings of the AAAI Conference on Artificial Intelligence, 34 (01), 1112–1119. doi: https://doi.org/10.1609/aaai.v34i01.5462
12. Reinders, H. J., Schoenmaker, D., van Dijk, M. (2023). A finance approach to climate stress testing. Journal of International Money and Finance, 131, 102797. doi: https://doi.org/10.1016/j.jimonfin.2022.102797
13. Emamat, M. S. M. M., Mota, C. M. de M., Mehregan, M. R., Sadeghi Moghadam, M. R., Nemery, P. (2022). Using ELECTRE-TRI and FlowSort methods in a stock portfolio selection context. Financial Innovation, 8 (1). doi: https://doi.org/10.1186/s40854-021-00318-1
14. Lukina, N. P., Nurgaleeva, L. V. (2005). Valuegical and ideological status of a network community in information social space: statement of a problem. Gumanitarnaya informatika.
15. Romanenkov, Y., Danova, M., Kashcheyeva, V., Bugaienko, O., Volk, M., Karminska-Bielobrova, M., Lobach, O. (2018). Complexification methods of interval forecast estimates in the problems on short-term prediction. Eastern-European Journal of Enterprise Technologies, 3 (3 (93)), 50–58. doi: https://doi.org/10.15587/1729-4061.2018.131939
16. Model velykykh danykh ta mashynnoho navchannia. Available at: https://business.diia.gov.ua/en/handbook/impact-investment/model-velikih-danih-ta-masinnogo-navcanna
17. Marbach, P., Tsitsiklis, J. N. (2001). Simulation-based optimization of Markov reward processes. IEEE Transactions on Automatic Control, 46 (2), 191–209. doi: https://doi.org/10.1109/9.905687
18. Raskin, L., Sukhomlyn, L., Sagaidachny, D., Korsun, R. (2021). Analysis of multi-threaded markov systems. Advanced Information Systems, 5 (4), 70–78. doi: https://doi.org/10.20998/2522-9052.2021.4.11
19. Raskin, L., Sira, O., Sukhomlyn, L., Parfeniuk, Y. (2021). Universal method for solving optimization problems under the conditions of uncertainty in the initial data. Eastern-European Journal of Enterprise Technologies, 1 (4 (109)), 46–53. doi: https://doi.org/10.15587/1729-4061.2021.225515

20. Raskin, L., Sira, O. (2016). Method of solving fuzzy problems of mathematical programming. Eastern-European Journal of Enterprise Technologies, 5 (4 (83)), 23–28. doi: https://doi.org/10.15587/1729-4061.2016.81292
21. Alibekov, E., Kubalík, J., Babuška, R. (2018). Policy derivation methods for critic-only reinforcement learning in continuous spaces. Engineering Applications of Artificial Intelligence, 69, 178–187. doi: https://doi.org/10.1016/j.engappai.2017.12.004
22. Semenov, S., Weilin, C., Zhang, L., Bulba, S. (2021). Automated penetration testing method using deep machine learning technology. Advanced Information Systems, 5 (3), 119–127. doi: https://doi.org/10.20998/2522-9052.2021.3.16
23. Zheng, L., Fiez, T., Alumbaugh, Z., Chasnov, B., Ratliff, L. J. (2022). Stackelberg Actor-Critic: Game-Theoretic Reinforcement Learning Algorithms. Proceedings of the AAAI Conference on Artificial Intelligence, 36 (8), 9217–9224. doi: https://doi.org/10.1609/aaai.v36i8.20908
24. Mnih, V. et al. (2016). Asynchronous methods for deep reinforcement learning. International conference on machine learning. doi: https://doi.org/10.48550/arXiv.1602.01783
25. Grondman, I., Busoniu, L., Lopes, G. A. D., Babuska, R. (2012). A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42 (6), 1291–1307. doi: https://doi.org/10.1109/tsmcc.2012.2218595
26. A faster, simpler approach to parallel Python. Available at: https://www.ray.io/ray-core