Computer Modelling and Investigation of Investment Portfolios

Nataliia Kuznietsova[a,b] , Oleksii Shevchuk[a]

[a] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", ave. Peremohy 37, Kyiv, 03056, Ukraine*

[b] *University Claude Bernard Lyon 1, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne cedex, France*

Abstract

In this paper the main methods and models for investment portfolio forming are studied. The novelty of the paper is development of the method for forming an investment portfolio based on artificial intelligence. It differs from classical models that study the economic metrics not separately for each company but in aggregate effect. Thus, it gives to the model an opportunity to consider the market as a whole and determine which stocks are overvalued and undervalued. The method solves the problem of forming an investment portfolio in two stages. At the first stage, the method finds the true price of the company's shares based on the company's information, its quarterly reports and daily financial metrics that describe the company's economic situation dynamics and its shares. At the second stage the found price is compared with the current market price and based on this information a decision is made to add shares to the portfolio, and with what percentage or discard them. It was also developed an information technology implemented the packages in Python and realized the data preparation and investigation modules, modules for formation an investment portfolio using various types of classical methods and based on them models as well as developed method based on artificial intelligence. An experimental study was performed on a sufficient number of real datasets and where the developed approach showed high enough prediction accuracy scores.

Keywords [1]

Investment analysis, computer modelling, investment portfolio, artificial intelligence-based models, Markowitz model, Bayesian networks, gradient boosting

1. Introduction

Nowadays, humanity moreover understands the proper of good knowledge of financial management importance. The main function of the management is to set financial assets in motion with the aim to them. Moreover, any country significantly depends on investment processes in the modern world.

The concept of financial investment dates to the time of ancient Babylon. The first written references to investing and competent financial management date back to those times [1]. Since that times, scientists have been looking for ways to improve investment efficiency by increasing returns and reducing the risk. One of the found solutions is the portfolio investment method.

The essence of portfolio investing is to improve investment opportunities by providing a set of investment objects with those investment qualities that cannot be achieved in the case of one single investment object, but are quite possible in the case of these objects combination. Thus, a new challenge for mathematicians and economists has emerged: finding the most optimal combination of investment portfolio object's.

An investment portfolio is a purposefully formed set of investments in investment objects that corresponds to a certain investment strategy of the investor by himself. It follows from this definition that the main goal of forming an investment portfolio is to ensure the developed investment policy

---

ICST-2023: Information Control Systems & Technologies, September 21-23, 2023, Odesa, Ukraine.
EMAIL: natalia-kpi@ukr.net (Kuznietsova N.); spichka.vin@gmail.com (Schevchuk O.)
ORCID: 0000-0002-1662-1974 (Kuznietsova N.); 0009-0005-7236-5417 (Schevchuk O.)

implementation by selecting the most effective and reliable investments. Depending on the direction of the chosen investment policy and the specifics of conducting investment activities, a system of specific goals is determined, the main and most common of which are: the capital growth maximization, the profit growth maximization, the investment risks minimization, ensuring the liquidity of the investment portfolio that meets the requirements.

The variety of types of investment objects, investment goals, their priority, as well as other conditions led to the creation of a huge list of types of investment portfolios characterized by a certain ratio of profitability and risk. The classification of investment portfolios by types of investment objects is primarily related to the direction and volumes of investment activity.

2. Problem statement

In this article the analysis of existing methods for an investment portfolio formation and investigation was carried out, their main advantages and disadvantages were identified. Some classical methods and models for investment portfolio formation such as the decision-making method with a fuzzy preference relation on a set of alternatives, Bayesian networks, Markowitz model and artificial intelligence-based method with gradient boosting models were discussed. The theoretical foundations of these methods were analyzed in accordance to their requirements, limitations and possibility for their implementation. It was also conducted the simulations based on these methods and comparative analysis on real data. The main aim in this article was to propose a new proprietary method for investment portfolio formation based on artificial intelligence methods. This AI method should offset main drawbacks of mentioned methods at all or at least partial to avoid limitations and give better results on the same dataset. It was also decided to develop an information technology that forms an investment portfolio using classical methods and a developed method based on artificial intelligence.

3. Methods and materials

Throughout the history of the investment portfolio forming task, a huge number of approaches and methods have been developed. A special contribution was made by such well-known economists as Sharpe and Markowitz. Also, a popular method for solving such problems is the Bayesian network method, which is based on Bayes' probabilistic theorem. With the advent of neural networks, many scientists are trying to solve some specific problems and tasks. The formation of an investment portfolio task was also the area of application and testing of different methods and approaches. Below will be discussed the most appropriate methods and models for solving the problem of investment strategy development with the application of classical and artificial intelligence methods and models.

3.1. Decision-making method

A forming investment portfolio process includes a key decision-making stage. It is very important that this stage is based on certain consequence of principles. Of course, the process of decision-making can be influenced by personal preferences or gut feelings, but the final decision should still be based on mathematical and statistical principles. This will help make the decision-making process somewhat idempotent [2]. This theory helps to formally describe the fuzzy concepts that people use to describe their desires, goals and perceptions of the system.

So now it is possible to determine the main advantage of the decision-making theory such as its simplicity and intuitiveness. Nevertheless, the factor of the existence of personal preferences rates usually cause inaccuracy in results and make the most significant disadvantage.

The decision-making process of choosing the most optimal alternative among a set of all alternatives can take place when different amounts of information about these alternatives are available. A universal way to describe input information is to write it in the form of a preference relation on a set of alternatives.

When modeling real systems, cases with clear relations of non-strict preference are rare. The case of fuzzy relations is more typical. Such relations are described by the membership function $\mu_R(x, y)$, which has the property of reflexivity. Given a fuzzy relation of non-strict preference R on X, we can unambiguously define three corresponding fuzzy relations [3]:

1. Fuzzy relation of indifference (equation 1)
2. Fuzzy equivalence relation (equation 2)
3. Fuzzy strict preference relation (equation 3)

$$\mu_R^J(x,y) = \max\left[\min\{1-\mu_R(x,y); 1-\mu_R(y,x)\}; \min\{\mu_R(x,y); \mu_R(y,x)\}\right] \quad (1)$$

$$\mu_R^e(x,y) = \min\{\mu_R(x,y); \mu_R(y,x)\} \quad (2)$$

$$\mu_R^S(x,y) = \begin{cases} \mu_R(x,y) - \mu_R(y,x), if\ \mu_R(x,y) > \mu_R(y,x) \\ 0, else \end{cases} \quad (3)$$

According to the definition of the intersection operation of fuzzy sets, the expression for the membership function of the set of non-dominated alternatives will be as follows:

$$\mu_R^{nd} = 1 - sup_{y \in X}[\mu_R^S(x,y)], x \in X \quad (4)$$

We must not forget that the main goal is to find the most profitable alternative x_0. To do this, the following property must be satisfied:

$$f_j(x_0) \geq f_j(y), \forall j = 1,..,m, \forall y \in X \quad (5)$$

If the condition is met, then the alternative x_0 is called Pareto-optimal and to solve the problem, we need to use another tool – convolution of multiple criteria into a scalar one. One of the most common ways of convolution is to use the intersection, which will give us the following:

$$\mu_{Q_1}(x,y) = min\{w_1\mu_1(x,y), \dots, w_m\mu_m(x,y)\} \quad (6)$$

where w_1,…, w_m are the weights of each criteria.

Another convolution important for finding optimal alternatives is the convolution of the original relations $\{R_J\}$, which is written as the sum:

$$\mu_{Q_2}(x,y) = \sum_{j=1}^{m} w_j\mu_j(x,y) \quad (7)$$

where $\sum_{j=1}^{m} w_j = 1, w_j \geq 0$.

Since the optimal values belong to the space of non-dominated alternatives, they will belong to the sets Q_1^{nd} and Q_2^{nd} obtained as a result of convolution. To obtain the optimal alternatives, it is necessary to find such elements from the joint set of non-dominated alternatives Q_1^{nd} and Q_2^{nd}. The non-dominance value will be maximal for optimal alternatives.

3.2. Markowitz model

The investment portfolio formation theory according to H. Markowitz [4] is based on the behavioral specifics of an investor who wants to invest its resources in companies with the lowest risk and receive certain dividends for this risk. Markowitz's approach assumes that an investor takes into account only two parameters: risk and return [5].

The Markowitz optimal portfolio model is based on the following principles:

1. An investor wants to maximize the income for a given risk level.
2. Investors always try to avoid risk. Between two assets with the same income, the one with the lower risk is chosen.
3. Risk is the uncertainty of a future outcome.
4. An investor's portfolio consists of all his assets and liabilities.
5. Investors make investment decisions based on expected returns and investment risk. The "usefulness of investments" is calculated by the formula:

$$U = E_R - A(E_Q)/2 \quad (8)$$

where A – the degree of risk aversion by the investor, E_R – expected income, E_Q – expected risk.

According to the Markowitz model, the portfolio income is the weighted average income on its components which is determined by the formula:

$$R_p = \sum_{i=1}^{N} W_i * r_i \quad (9)$$

where N is the number of stocks, r_i – income of a particular stock, W_i – the stock percentage.

The following equation is used to determine the riskiness of a portfolio:

$$Risk = \sqrt{\sum_{a=1}^{N}\sum_{b=1}^{N}(W_a\sigma_a W_b\sigma_b r_{ab})} \quad (10)$$

where W_i is the stock percentage, r_{ab} is a linear correlation coefficient, σ_a, σ_b are the risks of stocks a and b (standard deviation).

For Markowitz's model, the efficient set theorem is important, which states that an investor will choose his optimal portfolio among a set of portfolios, each of which will provide:

- maximum profitability for a given risk level;
- minimum risk for a given value of expected income.

The set of portfolios satisfying the theorem is called the efficient set or efficient frontier. Within this set or on the frontier are all portfolios that can be formed from a certain number of stocks. The effective set is the area in which the points are located, and the effective frontier is the line that graphically delineates this set (Figure 1).

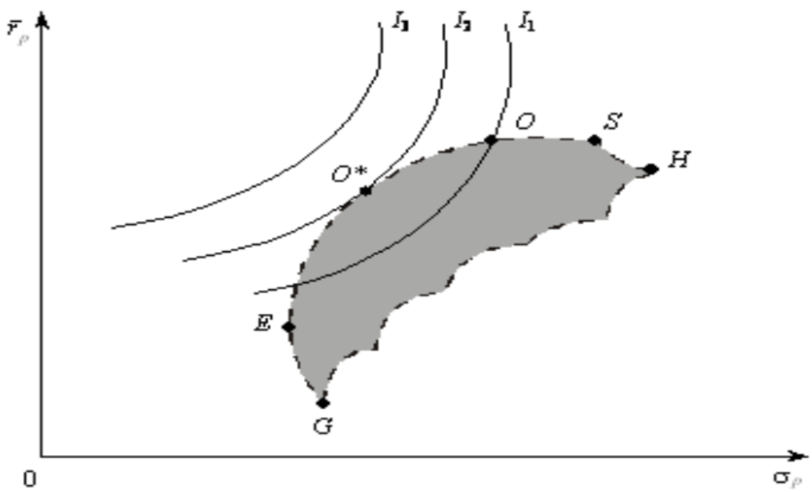

Figure 1: Effective set and frontier

The Pareto optimal portfolio for an investor, provided that his preferences and investment strategy are not taken into account, will be the portfolio that is located highest and to the left of all points in the set. On figure 1 such portfolio located at the point O^*.

The choice of investment portfolio objects depends not only on the objects themselves, but also on the investor's strategy and goals. Thus, the most obvious strategies are those of minimum risk and maximum return. Another way to form an investment portfolio is to maximize the Sharpe ratio [6].

This ratio helps to solve the problem of choosing between two investment objects by comparing their income with a risk-free object. The Sharpe ratio is calculated using the following equation:

$$S = \frac{R_p - R_f}{\sigma_p} \quad (11)$$

where R_p is a portfolio income, R_f is a risk-free object income, σ_p are the portfolio riskiness.

Thus, the Sharpe ratio will show the relative income on an asset per unit of risk. Maximizing this value will form a portfolio from the assets with the most "profitable" risk.

The main Markowitz model's advantage is that it is based on the economy theory and it was really efficient for solving economy tasks. But such model requires a big dataset of historical economy data which is not always possible to retrieve. Also for some tasks as the historical data could be not always in the same way as the relevant current financial processes. So the Markowitz model requires additional learning of actual data, otherwise it will lose accuracy.

3.3. Bayesian network

Bayesian networks are widely used for processing statistical data represented by time series. These networks play a special role in risk management. Bayesian networks help to establish cause-and-effect relationships between certain attributes and the conclusion obtained under such conditions [7].

A Bayesian network can be viewed as a model for representing the relationships between the vertices of an acyclic graph, which are represented as causal probability dependencies. Formally, a Bayesian network is a triple $N = < V, G, J >$, where V is a set of variables, G is a directed acyclic graph, and J is the joint probability distribution of the variables $V = \{ X_1, \ldots, X_n\}$. It should be mentioned, that the Markov condition is satisfied for the set of variables V, which means that each variable in the network is independent of all other variables except for the parental predecessors of this variable [5].

Bayesian networks are based on probability theory and, in particular, Bayes' theorem. In the general case where an event depends on several other events, Bayes' theorem is as follows:

$$p(D|S_1, \ldots, S_n) = \frac{p(D)p(S_1, \ldots, S_n|D)}{p(S_1, \ldots, S_n)} \tag{12}$$

where $p(D|S_1, \ldots, S_n)$ is the probability of an event D in the presence of a disturbances $S_1, \ldots, S_n$, $p(D)$ is the probability of the target event D (this value can be estimated using historical data), $p(S|D)$ is the probability of the reverse course of the event, the appearance of $S_1, \ldots, S_n$ in the presence of D, $p(S_1, \ldots, S_n)$ is the probability of occurrence of the events $S_1, \ldots, S_n$ (this value can also be estimated by analyzing historical data).

But this equation needs to be transformed to be used in a Bayesian network. Assuming that all events $S_1, \ldots, S_n$ are independent and D is known, the following equation will hold:

$$p(S_1, \ldots, S_n|D) = p(S_1|D)\ldots p(S_n|D) \tag{13}$$

With further normalization of equation 12, we can get rid of the denominator $p(S_1, \ldots, S_n)$, which will simplify the task of forming a conclusion. Thus, we obtain a generalized equation for drawing a conclusion by Bayes' theorem:

$$p(D|S_1, \ldots, S_n) = \alpha p(D)p(S_1|D)\ldots p(S_n|D) \tag{14}$$

The Bayessian model is really good for classification and it can give a probability of each event, but it requires also a big dataset, and the model based on the Bayesian network grows with incresing number of features. Another drowback is that Bayesian model is not able to predict an accurate value but just classify the sample to a group or gives the probability of the appearance of some fact. So for financial tasks you need carefully to make a problem statement as well the main "probability" evidence which you are going to forecast.

3.4. Artificial intelligence-based models

There are numerous of different artificial intelligence models, ranging from various variations of gradient boosting to full-fledged neural networks. Often, a single model is not enough to solve a certain task. To solve this problem, an ensemble of models is used, i.e., several machine learning algorithms assembled into a single one. This approach is often used to enhance the positive qualities of individual algorithms, which may be weak on their own, but show excellent results in a group. When using ensemble methods, algorithms are trained simultaneously and can correct each other's mistakes. Model ensembles are usually built on the basis of a decision tree model. Trees are added one at a time to the ensemble and trained to mutually correct prediction errors made by previous models. This type of ensemble is called boosting [8].

Gradient boosting is one of the classic models of artificial intelligence (AI). It also belongs to the ensemble models of artificial intelligence, i.e., this AI model will consist of several models [9, 10].

Consider the task of recognizing objects in a multidimensional space X with a label space Y. Suppose we are given a training set $\{x_i\}_{i=1}^{N}$, where $x_i \in X$, and the true labels $\{y_i\}_{i=1}^{N}$, $y_i \in Y$, of each object from the set. The learning task is to use the training set to find the approximating function $\widehat{F}(x)$ to the function $F(x)$ which minimizes the expected value of some given loss function according to the formula:

$$\hat{F}(x) = argmin_F \mathrm{E}_{x,y}[L(y, F(x))] \quad (15)$$

where $L(y, F(x))$ is a loss function.

The gradient boosting method searches for an approximating function $\hat{F}(x)$ as a weighted sum of functions $h(x)$ and some class H, which are considered as weak models. With this in mind, the equation (15) can be represented:

$$\hat{F}(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + const \quad (16)$$

where M is a number of weak models, γ_i is a weighted coefficient and $h_i(x)$ is a weak model function.

Thus, the output of the learning algorithm is a set of M decision trees, and to make a prediction, i.e., to determine the output y for a new object x, we should calculate the sum with the equation:

$$y = T_0 + v * \sum_{m=1}^{M} T_m(x) \quad (17)$$

where T_0 is the first decision tree, v is the scaling factor, M is the total number of constructed decision trees, T_m the m-th decision tree.

The final classifier is represented as a linear combination of classifiers. Finding the optimal values of the coefficients of this linear combination is a rather time-consuming task, so gradient boosting uses a greedy algorithm for gradually adding classifiers [10, 11].

The most significant advantages of AI-based models are their flexibility and possibility to be specialized for the certain task. It makes results more accurate then with other models. Another advantages goes from using ensembles of the models. It gives opportunity to resolve tasks with a lack of data in the dataset. So, the approach based on using of AI models seems to us more perspective, modern and relevant for using to practical financial tasks. In this paper we will focus on development of the AI models for forming an investment portfolio. Here we propose the new method based on AI which differs from classical that study the economic metrics not separately for each company but their aggregate effect. Consequently, it gives to the model an opportunity to consider the whole financial market and to classify which companies' stocks are overvalued and undervalued and to choose the most attractive and valuable companies for investment.

4. Modeling and simulation of the practical task

It was decided to build the following models for testing the forming investment portfolios effectiveness in different ways, such as:

- decision-making model with fuzzy preference relation on a set of alternatives;
- Bayesian network model;
- Markowitz model;
- artificial intelligence model.

4.1. Data sample description

The key data for forming an investment portfolio task is the general information about the company, as well as its financial position and the dynamics of its development. Thus, three types of data were used: general information about companies, quarterly companies' reports, daily information about companies and their stocks. Basic data refers to all information about a company that generally does not change over time. This includes information about the industry in which the company specializes, the sector in which the company operates, etc. (Table 1).

Table 1
Base data snippet

Ticker	Sector	Industry	Currency
NFLX	Consumer Cyclical	Entertainment	USD
NVDA	Technology	Semiconductors	USD
SONY	Technology	Consumer Electronics	JPY
TM	Consumer Cyclical	Auto Manufactures	JPY
TSLA	Consumer Cyclical	Auto Manufactures	USD

Quarterly reports include a variety of information on the company's development dynamics, as well as typical metrics used to assess the company's financial performance. Such data include the amount of debt, the amount of the company's profit, information on annual stockholder dividends, profit before tax, etc. (Table 2).

Table 2
Quarterly data snippet

Ticker	Revenue	Netinc	NCF	EBITDA	FCF	Current Ratio
NVDA	76430000000	3003000000	702000000	3235000000	2760000000	6.650
NVDA	71030000000	2464000000	-4340000000	2998000000	1297000000	7.145
NVDA	65070000000	2374000000	4650000000	2740000000	2499000000	5.802
NVDA	56610000000	1912000000	1310000000	2378000000	1576000000	4.527

The last type of information is daily data, which describes the daily dynamics of the financial attractiveness of a company and its stocks. Such data includes information about the company's capitalization and the closing price of the companies' stocks (Table 3).

Table 3
Daily data snippet

Ticker	Date	Marketcap	Close
NVDA	2022-04-29	465529.7	185.47
NVDA	2022-04-28	496528.2	197.82
NVDA	2022-04-27	462216.5	184.15
NVDA	2022-04-26	471578.8	187.88
NVDA	2022-04-25	499540.2	199.02

The entire dataset includes a five-year period. For model training, data up to 01-07-2022 is used. The validation data ends on 28-09-2022. The number of companies used to train the models is significant and exceeds 500. But the number of companies from which the portfolio will be formed is limited. For each of four different industries six were selected:

- car industry tickers: NKLA, TSLA, RACE, STLA, F, TM;
- software development tickers: MSFT, ADBE, ORCL, PAYO, BB, DBX;
- electrical appliances' industry: AAPL, NVDA, AMD, SONY, KOSS, BOX;
- entertainment – NFLX, DIS, IMAX, WWE, CNK, WMG.

4.2. Development of the Artificial Intelligence-based model

The decision-making model, that was built to select the optimal investment objects for an investment portfolio, received input features that describe the company's financial position and growth dynamics. The most important features were the industry in which the company operates, changes in the stock price, company capitalization and the debt amount. The data was taken for the last quarter and the difference data was calculated as the difference between the last and the penultimate quarter.

The model output was a metric of each alternative importance, which ranges from 0 to 1. The higher metric value corresponds to the higher priority of choosing this alternative. The number of companies to be added to the portfolio was set by a user. The stocks' percentage in each company was calculated by the formula:

$$percent\ _i = \frac{metric_i}{\sum_{i=1}^{N} metric_i} * 100 \tag{18}$$

where $metric_i$ is a metric returned by the model for the i-th company, N is a number of companies in the portfolio and in experiments this number was chosen as 7.

The Bayesian network model receives various data about the companies from which the portfolio is planned to be formed, as well as their financial indicators' dynamics. As an output, the model returns the probability that the company's stock will grow in value. Only those companies will be selected for the portfolio whose growth probability exceeds a certain cut-off point set by the investor. The percentage of each investment object is determined by the following formula:

$$percent_{\,i} = \frac{prob_i}{\sum_{i=1}^{N} prob_i} * 100 \qquad (19)$$

where $prob_i$ is the probability of increase in the stock price of the i-th company, N is a number of companies whose stock price growth probability exceeded the cut-off point.

The Markowitz method was implemented by searching through one hundred million potential portfolios. The portfolios from the portfolio effective set that were optimal according to the risk minimization strategy and the Sharpe ratio maximization strategy were selected.

Specific data pre-processing for the artificial intelligence-based model was carried out. The model received all types of data: basic, daily, and quarterly. Basic text data is converted into numeric data. For daily and quarterly data, statistics were obtained for each metric for different specified intervals, such as standard deviation, an average value for the period, and others.

It was developed the method based on artificial intelligence for building investor's portfolio which works in two stages. In the first stage, it uses an ensemble of gradient boosting models to determine the "fair/true" share price. In the second stage it uses the equation (20) to determine whether the company is undervalued or overvalued.

$$Coef = \frac{\text{"Fair"}\ Stock\ Price}{Real\ Stock\ Price} \qquad (20)$$

where "Fair" $Stock\ Price$ is the price determined by the gradient boosting method and $Real\ Stock\ Price$ is the price from real dataset. If the $Coef < 1$ it means that the company is undervalued and in next period its shares will increase, so it is recommended to add this company in investor's portfolio. If the $Coef > 1$ it means that the company is overvalued and in next period the shares will fall and it is not recommended to investors to focus on this company and to include it in portfolio.

Companies for which this ratio exceeds a certain cut-off point determined by analysts beforehand will also be rejected as erroneous. The share of each investee is calculated as the ratio of the investee's coefficient to the sum of all selected investees coefficients.

5. Results

5.1. Decision-making model

The first built model was a decision-making model with a fuzzy preference relation on a set of alternatives. An investment portfolio with seven investment objects was formed (Table 4).

Table 4
Decision-making model portfolio

Company	Price diff	Metric	Percent	Income
TSLA	-64.24	1	21.74	-1396.52
NKLA	0.29	0.75	16.30	4.73
WWE	12.68	0.7	15.22	192.96
BOX	-1.00	0.6	13.04	-13.04
STLA	-4.77	0.55	11.96	-57.03
BB	-0.76	0.5	10.87	-8.26
AMD	-1.33	0.5	10.87	-14.46

The total portfolio income is -1291.63. This indicates unsatisfactory model results. Most of the investment targets selected by the model lost their price. TSLA stocks suffered the biggest losses. Its stocks fell the most out of the entire set of alternatives, and the model chose these stocks as the most desirable. On the positive side, we can point out the choice of WWE as a target. The stocks of this company grew the most out of the entire set, but it is not enough to compensate the losses of this portfolio.

Thus, the decision-making model did not produce a satisfactory result. It selected only those investment targets that the investor liked, with little regard for financial feasibility. The assignment of weights to each criterion has a negative impact on the model's efficiency, as they are based on certain subjective judgments rather than analytical decisions.

5.2. Bayesian network model

The Bayesian network model receives various companies input data from which the portfolio is planned to be formed, as well as their financial indicators dynamics. At the output model returns the probability that the company's shares will increase in price. Only those companies whose growth probability will exceed a certain cut-off limit set by the investor will be selected for the portfolio. Daily data were converted to quarterly and combined with preliminary quarterly data for Bayesian network model. Basic information about each company's sector and industry was also added to them. The first difference for numerical data was taken in order to form the task of determining the share price's growth or fall based on the input data. The value 0.55 was chosen as the cut-off, i.e. all companies whose stock growth probability exceeds 55%. Received results for Bayesian modeling are illustrated in the Table 5.

Table 5
Bayesian model portfolio

Company	Price diff	Probability	Percent	Income
MSFT	6.19	56.14	11.41	70.63
ADBE	-87.1	56.12	11.41	-993.42
ORCL	8.36	56.14	11.41	95.38
PAYO	2.08	56.12	11.41	23.72
BB	-0.76	56.12	11.41	-8.67
DBX	2.41	56.12	11.41	27.49
SONY	2.76	99.17	20.15	55.62
BOX	-1	56.12	11.41	-11.41

The total portfolio income is -740.65. The resulting portfolio created by the Bayesian network model was also found to be unprofitable. Compared to the previous portfolio, created by decision-making model, it has some advantages, in addition to a smaller loss. This includes numerous of profitable stocks. Only three of the eight investment objects fell in value, and unfortunately, the fall of one of them was significant. All the other objects were characterized by price growth. In addition, if the cut-off value was increased, the portfolio risk would decrease, and, in this particular case, there would be only one SONY investment target that had an increase in price.

5.3. Markowitz model

The Markowitz model was implemented by sifting through one hundred million portfolios. As a result, the cloud of portfolios in the Figure 2 was obtained. Two investment strategies for the Markowitz model were considered: the minimum risk strategy and the maximum Sharpe ratio strategy. On the Figure 2, red point represents a portfolio for minimum risk strategy, and green – for maximum Sharp ratio strategy. The results of each strategy are presented in Tables 6 and 7.

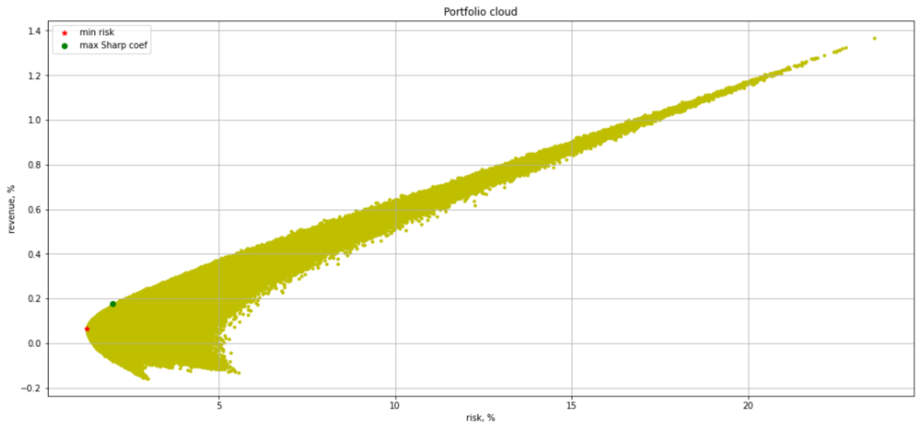

Figure 2: Cloud of portfolios according to the Markowitz model

Table 6
Minimum risk Markowitz model portfolio

Company	Price diff	Percent	Income
TSLA	-64.24	1.4	-89.94
STLA	-4.77	2.3	-10.97
TM	-8.15	33.3	-271.40
MSFT	6.19	2.8	17.33
ADBE	-87.1	1.4	-121.94
ORCL	8.36	16.9	141.28
DBX	2.41	3.4	8.19
NVDA	-37.17	0.9	-33.45
AMD	-1.33	2.5	-3.33
SONY	2.76	3	8.28
BOX	-1	2.6	-2.6
NFLX	5.66	1.1	6.23
DIS	7.22	2.6	18.77
IMAX	1.83	1.1	2.01
WWE	12.68	16.4	207.95
CNK	1.57	1.3	2.04
WMG	3.72	3.3	12.28

Table 7
Max Sharp ratio Markowitz model portfolio

Company	Price diff	Percent	Income
TSLA	-64.24	3.9	-250.54
F	-0.62	15.5	-9.61
TM	-8.15	2.7	-22.01
ADBE	-87.1	2.1	-181.91
ORCL	8.36	1.3	10.87
AAPL	4.86	1.6	7.78
NVDA	-37.17	2.7	-100.36
AMD	-1.33	1.6	-2.13
KOSS	1.75	4.5	7.88
BOX	-1	5.9	-5.9
DIS	7.22	1.3	9.39
IMAX	1.83	1.3	2.38
WWE	12.68	48.9	620.05
CNK	1.57	3.8	5.97

Table 6 shows the most significant elements of the Markowitz model of minimal risk portfolio. The total minimum risk Markowitz model portfolio income is -105.67. The portfolio showed much better result than the previous ones, but it is still unprofitable. The benefits of the portfolio are that the most unprofitable investment objects such as ADBE, TSLA, NVDA are taken with a very low percentage. In addition, the most profitable stocks, such as WWE and ORCL, account for almost a third part of the portfolio. The biggest portfolio loss was caused by buying TM shares with a significant percentage. Usually, these stocks are relatively stable, but this time they fell sharply during the quarter.

It is also significant to mention that the market tended to fall during 2022 due to geopolitical factors. If in the sample were investment objects that had risen approximately as much as TSLA or ADBE fell, the portfolio could have made zero losses or even a profit.

The total maximum Sharp ratio Markowitz model portfolio income is 96.75. The portfolio was profitable, but this profitability was achieved through the adventurous purchase of WWE shares, which make up almost half of the total portfolio. In addition, the shares of TSLA, ADBE, and NVDA increased, which reduced the portfolio potential returns.

5.4. Artificial intelligence-based model

The portfolio formed by the artificial intelligence-based model showed the best results (Table 8). The total income is 195.93. In addition, all the most unprofitable stocks, such as TSLA and ADBE, were rejected, while the most profitable ones, such as ORCL or WWE, were added to the portfolio. It is also worth noting that the portfolio is quite balanced. The share of each investment objects is fairly equal, which makes the portfolio relatively stable, unlike the portfolio of the Markowitz maximum Sharpe ratio model. The main drawback of this model was training time. It took much more time and processing resources than any previous model.

Table 8
AI model portfolio

Company	Price diff	Coeff	Percent	Income
NKLA	0.29	1.52	8.96	2.60
RACE	-0.71	1.03	6.08	-4.32
STLA	-4.77	1.82	10.72	-51.13
ORCL	8.36	1.61	9.50	79.44
PAYO	2.08	2.34	13.80	29.71
AMD	-1.33	1.67	9.81	-13.05
SONY	2.76	1.63	9.61	26.53
BOX	-1.00	1.46	8.62	-8.62
WWE	12.68	1.22	7.20	91.35

6. Discussion

After obtaining the simulation results, it is possible to discuss the effectiveness of the classical approaches as well as the developed method for investment portfolio formation. It is obvious that the results of decision-making methods with a fuzzy preference ratio on a set of alternatives and Bayesian networks showed too bad results, so it was decided not to add them to the final comparative Table 9.

The comparison table clearly shows that the developed artificial intelligence-based method for investment portfolio formation rejected the most unprofitable objects that were included in the portfolios created by the Markowitz models of minimum risk and maximum Sharpe ratio. The following advantage of the developed method is also clearly visible – the objects of its portfolio are maximally equivalent, which is the most weighted. Even in a minimal risks Markowitz portfolio there is an asset like TM, that clearly outperforms all other assets in the portfolio in percentage terms. The main disadvantage of the developed method is its demand for computational capabilities and input data volumes. The method uses a huge amount of data about different companies in order to determine the true shares price. Accordingly, it is needed to spend a lot of time on this. The same Markowitz method will run much faster because it doesn't need to process as much data. On the other hand, a trained model can form a portfolio from a new set of companies that could be included into the model, whereas a Markowitz model needs to start forming a portfolio from beginning with new data. In this case, the developed artificial intelligence-based method will work faster.

Table 9
Comparison of portfolios

Company	Price difference	Markowitz (min. diff.)		Markowitz (max. Sharpe coefficient)		The developed artificial intelligence method	
		Percentage	Profit	Percentage	Profit	Percentage	Profit
TSLA	-64.24	1.4	-89.94	3.9	-250.54	-	-
F	-0.62	0.6	-0.37	15.5	-9.61	-	-
NKLA	0.29	0.5	0.15	0.2	0.06	8.96	2.60
RACE	-0.71	0.5	-0.36	0.5	-0.36	6.08	-4.31
STLA	-4.77	2.3	-10.97	0.2	-0.95	10.72	-51.13
TM	-8.15	33.3	-271.40	2.7	-22.01	-	-
MSFT	6.19	2.8	17.33	0.7	4.33	-	-
ADBE	-87.1	1.4	-121.94	2.1	-181.91	-	-
ORCL	8.36	16.9	141.28	1.3	10.87	9.50	79.43
PAYO	2.08	0.8	1.66	0.3	0.62	13.80	29.71
DBX	2.41	3.4	8.19	0.5	1.21	-	-
AAPL	4.86	0.5	2.43	1.6	7.78	-	-
NVDA	-37.17	0.9	-33.45	2.7	-100.36	-	-
AMD	-1.33	2.5	-3.33	1.6	-2.13	9.81	-13.05
SONY	2.76	3.0	8.28	0.1	0.28	9.61	26.53
KOSS	1.75	0.3	0.53	4.5	7.88	-	-
BOX	-1.00	2.6	-2.6	5.9	-5.9	8.62	-8.62
NFLX	5.66	1.1	6.23	0.1	0.57	-	-
DIS	7.22	2.6	18.77	1.3	9.39	-	-
IMAX	1.83	1.1	2.01	1.3	2.38	-	-
WWE	12.68	16.4	207.95	48.9	620.05	7.20	91.35
CNK	1.57	1.3	2.04	3.8	5.97	-	-
WMG	3.72	3.3	12.28	0.1	0.37	-	-
Total profit		-105.67		96.75		195.93	

It was also developed the information technology (IT) using Python 3.10.4 and Anaconda 4.9.2 and the main libraries such as Numpy, Pandas, Sklearn. A new software product for the investment portfolio formation using various types of methods and appropriate models was developed. The information technology has a complex structure and consists of a large number of applications and modules. In particular, separate modules for preliminary data preparation and processing (download.ipynb, data_loaders.ipynb, utils.ipynb, features.ipynb and targets.ipynb) have been implemented. Several modules are used to form models of the developed method based on artificial intelligence: ai_model.ipynb metrics.ipynb, models.ipynb, pipeline.ipynb. The markovitz_model.ipynb module is used to build and conduct experiments based on the Markowitz model of minimum risk and maximum Sharpe ratio. The module bayesian_model.ipynb presents the results of modeling on Bayesian networks, and multicriteria_model.ipynb – the construction of a decision-making model with a fuzzy preference ratio on a set of non-dominated alternatives. The disadvantage of the IT is its significant consumption of the users' device resources, in particular for Bayesian methods and the developed artificial intelligence-based decision-making method. This problem could be solved by optimizing the program as well as by uploading the IT to cloud services. It is also advisable to use parallel calculations.

7. Conclusions

The main existed methods for forming an investment portfolio, from the simplest classical to the most modern methods of AI were considered. Based on the most relevant to this task methods, the models for investment portfolio formation were built and compared. Obtained result showed that the portfolios formed by the decision-making method and the Bayesian network method showed unsatisfactory results. For the decision-making method, the explanation for such unsatisfactory results is that it does not analyze the financial market, but simply helps in decision-making. The results on investment modelling by Markowitz model were much better. This explains its popularity among analysts and investors. In addition, Markowitz model gives the option to calculate the portfolio risk.

As a result, a method of forming an investment portfolio based on artificial intelligence was developed and compared with classical methods. Then it was tested on the real data and received results of comparison showed that the developed method has a higher efficiency than classical methods. The resulting investment portfolio produced the highest return among the other portfolios and was also balanced, which shows its reliability. Results could also be improved by using even more numbers of different metrics and increasing the size of the dataset. The advantage of the artificial intelligence-based method is that it analyzes the market as a whole and can compare companies with each other. In comparison the Markowitz model only analyzes the time series of share prices.

As it is shown by experiments, the proposed artificial intelligence-based method should be also improved in future research. The first way to improve modeling results is to expand the dataset. This can be done by extending the period for which the data was collected, for example, for 10 years. It should be also increased the number of metrics themselves, add other financial indicators [12] that describe the company's position, in addition to those mentioned earlier. In addition it could be also proposed to use other methods for investment portfolio formation. For example, neural networks, in particular those that work with fuzzy data. The approach of detecting shares and companies for investing could also be adapted by the investor's behavior and its attitude to risks.

8. References

[1] F. Ponziani, The evolution of investing from Babylon to MELD, 2022. URL: https://www.meld.com/blog/the-evolution-of-investing-from-babylon-to-meld.

[2] M. Voskoglou. Fuzzy sets, Fuzzy Logic and Their Applications, 2021.

[3] Yu.P. Zaichenko, Operations research, Slovo Publishing House, 2006.

[4] H. Markowitz, Portfolio selection: Efficient Diversification of Investments, 1991.

[5] P. Bidyuk, S. Polozhaenko, N. Kuznietsova, L. Levenchuk, Probabilistic Data Analysis in Non-Stationary Processes Forecasting, in: Proc. 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, pp. 297-302, doi: 10.1109/SAIC51296.2020.9239200.

[6] S. E. Paw. The Sharpe Ratio: Statistics and Application, ed. 1, 2021.

[7] J. Pearl, S. Russell, Bayesian Networks, in: M.A. Arbib (Ed.), Handbook of Brain Theory and Neural Networks, Cambridge, MA: MIT Press, 2003.

[8] Freund Y., Schapire R., Boosting: Foundations and Algorithms (Adaptive Computation and Machine Learning Series), 2014.

[9] C. Zhang, Y. Ma Ensemble Machine Learning: Methods and Applications, 2012.

[10] Mason L., Baxter J., Bartlett P., Frean M., Boosting Algorithms as Gradient Descent, Neural Information Processing Systems, 12 (2000) 512-518.

[11] Galkina S., Optimal Portfolio Selection Using Machine Learning Techniques, International Journal of Open Information Technologies, 2 6 (2014) 14-19.

[12] P.I. Bidyuk, N. V. Kuznietsova, Forecasting the volatility of financial processes with conditional variance models, Journal of Automation and Information Sciences, 46 10 (2014) 11-19.