

Модификация алгоритма «Случайный лес» для выбора инвестиционных инструментов

А.В. Зиненко

Сибирский федеральный университет,
660041, г. Красноярск, просп. Свободный 79, Россия

Резюме. Цель. Целью исследования является повышение эффективности инвестиционных решений путем разработки алгоритма машинного обучения «Модифицированный случайный лес» для формирования инвестиционного портфеля. **Метод.** Предложен алгоритм бинарной классификации «Модифицированный случайный лес», основанный на алгоритмах машинного обучения «Дерево решений» и «Случайный лес». На первом этапе алгоритм строит дерево решений, основанное на прогнозах наивным методом и методом ARIMA, на втором этапе – формирует «лес» деревьев из случайных подвыборок. **Результат.** Алгоритм протестирован на разных временных промежутках на инструментах российского индекса Мосбиржи и американского индекса NASDAQ. Для оценки точности работы алгоритма были выбраны метрики Precision, Recall, Accuracy. Для сравнения был произведен отбор акций в портфель при тех же условиях классическим методом задачи Марковитца, результаты которого по соответствующим метрикам оказались несколько слабее. Предложенный алгоритм занимает существенно меньше времени работы, поскольку не решает задачу оптимизации при формировании портфеля. **Вывод.** Разработанный алгоритм машинного обучения «Модифицированный случайный лес», основанный на методах прогнозирования Arima и наивный прогноз, позволяет решить задачу повышения эффективности управления инвестиционным портфелем, а также в силу своей бинарности может быть использован не только в финансовой сфере, но и для задачи классификации любых других прогнозируемых объектов.

Ключевые слова: финансовые рынки, финансовый анализ, ARIMA-модель, дерево решений, метод случайного леса

Для цитирования: А.В. Зиненко. Модификация алгоритма «Случайный лес» для выбора инвестиционных инструментов. Вестник Дагестанского государственного технического университета. Технические науки. 2025;52(3):38-48. DOI:10.21822/2073-6185-2025-52-3-38-48

Modification of the Random Forest Algorithm for Selecting Investment Instruments

A.V. Zinenko

Siberian Federal University,
79 Svobodny Ave., Krasnoyarsk 660041, Russia

Abstract. Objective. The aim of the study is to improve the efficiency of investment decisions by developing a machine learning algorithm "Modified Random Forest" for forming an investment portfolio. **Method.** The article proposes a binary classification algorithm "Modified Random Forest" based on the machine learning algorithms "Decision Tree" and "Random Forest". At the first stage, the algorithm builds a decision tree based on forecasts using the naive method and the ARIMA method, and at the second stage, it forms a "forest" of trees from random subsamples. **Result.** The algorithm was tested on different time intervals on the instruments of the Russian Moscow Exchange index and the American NASDAQ index. To assess the accuracy of the algorithm, the Precision, Recall, and Accuracy metrics were selected. For comparison, shares were selected for the portfolio under the same conditions using the classical Markowitz problem method, the results of which for the corresponding metrics were somewhat weaker. At the same

time, the proposed algorithm takes significantly less time to run, since it does not solve the optimization problem when forming a portfolio. **Conclusion.** The developed machine learning algorithm "Modified random forest", based on the Arima forecasting methods and naive forecast allows solving the problem of increasing the efficiency of investment portfolio management, and due to its binary nature, it can be used not only in the financial sector, but also for the task of classifying any other predicted objects.

Keywords: financial markets, financial analyses, optimal investment portfolio, ARIMA model, decision tree, random forest method

For citation: A.V. Zinenko. Modification of the Random Forest Algorithm for Selecting Investment Instruments. Herald of Daghestan State Technical University. Technical Sciences. 2025;52(3):38-48. (In Russ.) DOI:10.21822/2073-6185-2025-52-3-38-48

Введение. Задача формирования инвестиционного портфеля является одной из ключевых областей инвестиционного анализа. Целью как частного, так и институционального инвестора является отдача от вложенных денежных средств с минимальным риском. Мерой риска выступает неопределенность; математически риск инвестиций оценивается как отклонение доходности от ожидаемого значения. Ожидаемое значение доходности рассчитывается, как математическое ожидание доходностей за прошлые периоды, а риск – как дисперсия или стандартное отклонение. Впервые подход, представляющий динамику цен финансовых инструментов как случайный процесс с математическим ожиданием и дисперсией, был предложен Луи Башелье в 1900 году [1]. Диссертационная работа «Теория спекуляций», посвященная французским бессрочным облигациям, не имела успеха у оппонентов, поэтому его теория на долгое время была проигнорирована.

Только спустя 50 лет после защиты диссертации «Теория спекуляций» Луи Башелье его подход к оценке динамики финансовых рынков стал повсеместно распространен и смог называться классическим. В 1952 году американский экономист Гарри Марковитц. Опубликовал свою работу «Выбор портфеля», дав подходу Башелье новую жизнь.

Аналогично Башелье, Марковитц определял доходность ценной бумаги как математическое ожидание, а риск – как стандартное отклонение или дисперсию. Имея n ценных бумаг, каждая из которых имеет свою доходность и свой риск, инвестор может составить из них портфель уже с новыми риском и доходностью.

Оптимальным портфелем, согласно Марковитцу, является такой портфель, который при заданной доходности имеет минимальный риск (прямая задача) либо при заданном риске имеет максимальную доходность (обратная задача). Таким образом, решением задачи Марковитца является нахождение таких долей активов, при которых инвестиционный портфель является оптимальным. Модель Марковитца представлена в соответствии с (1).

$$\begin{aligned} V_p &= X^T V X \rightarrow \min, \\ M^T X &= m_p, \\ I^T X &= 1, \end{aligned} \quad (1)$$

где V_p – целевая функция (дисперсия портфеля); V – матрица ковариаций (диагональные элементы этой матрицы являются дисперсиями); X – матрица-столбец долей активов; M – матрица-столбец ожидаемых доходностей активов, входящих в портфель; I – единичная матрица – столбец; m_p – заданное желаемое значение доходности портфеля.

Данная задача решается стандартным методом математического программирования. Находятся и приравниваются к нулю частные производные функции Лагранжа. В готовом виде решение представлено в работе [2] и выглядит следующим образом (2).

$$\begin{aligned} X^* &= V^{-1} \frac{m_p (I J_{12} - M J_1) + M J_{12} - I J_2}{J_{12}^2 - J_1 J_2} \\ J_{12} &= I^T V^{-1} M, J_1 = I^T V^{-1} I, J_2 = M^T V^{-1} M \end{aligned} \quad (2)$$

В настоящее время существует немало модификаций модели Марковитца. Модель Тобина [3] добавляет в портфель безрисковый актив. Также, убрав ограничение по неотрицательности долей, возможно добавлять в портфель короткие позиции.

При своей распространенности модель Марковитца имеет некоторые недостатки. Прежде всего, стандартный метод оптимизации выбирает из всего множества активов только небольшое количество, удовлетворяющее поставленным условиям. Таким образом, анализируя индекс, состоящий из сотен активов, модель может выбрать только единицы, что является деструктивным фактором с точки зрения диверсификации. Другой недостаток модели – это сильная чувствительность к начальным условиям [3]. При небольшом изменении дисперсии одного из активов, структура портфеля может существенно поменяться.

В 1964 году ученик Марковитца Уильям Шарп разработал модель ценообразования рынка капиталов (САРМ), основанную на предпосылках Марковитца и Башелье. Не менее интересной является его рыночная модель, разработанная несколько ранее. Комбинация рыночной модели и модели САРМ дает альтернативный задаче Марковитца метод формирования оптимального инвестиционного портфеля [4]. Основная идея модели Шарпа – это разделение доходности и риска каждого инструмента на две компоненты: общерыночную и специфическую.

Таким образом, доходность актива представляется в виде уравнения линейной регрессии, где фактором выступает доходность рыночного портфеля, сдвигом – специфическая доходность, а коэффициентом чувствительности к рынку – бета-коэффициент. Риск в модели Шарпа также делится на рыночную и специфическую компоненты; от специфической компоненты можно избавиться, составив рыночный портфель. Модель оптимального портфеля по Шарпу представлена в соответствии с (3).

$$\begin{aligned} \sigma_p^2 &= \sum_{i=1}^n (x_i \beta_i)^2 \sigma_m^2 + \sum_{i=1}^n x_i^2 \delta_i^2 \rightarrow \min \\ r_p &= r_f + (\sum_{i=1}^n x_i \beta_i) r_m = m \\ \sum_{i=1}^n x_i &= 1, \end{aligned} \quad (3)$$

где σ_p^2 – риск портфеля; x_i – доля i -й акции в портфеле; β_i – бета-коэффициент i -й акции; σ_m^2 – рыночный риск; δ_i^2 – специфический риск i -й акции (дисперсия остатков регрессионной модели); r_p – доходность портфеля; r_f – безрисковая ставка; r_m – рыночная доходность; m – заданная доходность портфеля.

К недостаткам модели Шарпа можно отнести, во-первых, длительные вычисления, так как для каждого рассматриваемого актива необходимо строить уравнение регрессии с рынком. Также сам подход предполагает сильную корреляцию акций с рынком; соответственно, если рынок падает, то портфельный аналитик ничего не сможет сделать.

Постановка задачи. Нами предлагается отбор акций для портфеля без оптимизации, другими словами бинарная классификация – либо включаем акцию в портфель, либо не включаем. Затем, к отобранным акциям возможно применить модификации рассмотренных классических оптимизационных моделей.

Задача бинарного выбора акций для инвестиционного портфеля встречается в источниках достаточно редко [5] и представлена в виде задачи целочисленного программирования «задача о рюкзаке». Формальное описание задачи о рюкзаке для формирования портфеля акций представлено в соответствии с (4).

$$\begin{aligned} R &= \sum_{i=1}^n x_i (\hat{p}_i - p_i) \rightarrow \max \\ \sum_{i=1}^n x_i p_i &\leq I, \end{aligned} \quad (4)$$

где R – целевая функция – доходность портфеля; n – количество анализируемых акций; x_i – бинарная переменная, принимает значение 0 или 1; $\hat{p}_i$ – прогнозируемая стоимость акции; p_i – текущая стоимость акции; I – располагаемая сумма инвестора.

Такая задача может быть решена методом ветвей и границ, либо методом динамического программирования [5]. При достаточно простой формулировке «задача о рюкзаке» успешно определяет акции, которые можно включить в портфель. В работе авторов [6] представлена модификация задачи о рюкзаке, устраняющая необходимость включать в расчеты располагаемый инвестором доход. Данная модель показала хороший баланс

«Простота / Результативность», однако имеет смысл разработка более сложных и при этом более результативных моделей.

Существенной проблемой всех рассмотренных моделей является их неприменимость в стадии падения рынка. К тому же классические модели предполагают случайный характер биржевых котировок, тогда как по факту случайному закону подчиняются не все биржевые временные ряды.

Созданный нами алгоритм «Модифицированный случайный лес» позволяет решить обе проблемы. Метрики эффективности и функция ошибки, на которых обучается модель не основываются на доходности активов. Использование модели ARIMA решает вторую проблему, так как предполагает взятие разниц первого порядка, что приводит неслучайный временной ряд к стационарному виду.

Методы исследования. Разработанный метод формирования портфеля инвестиций «Модифицированный случайный лес» включает следующие этапы:

- 1) построение прогнозов методами ARIMA и Наивный прогноз;
- 2) построение дерева решений на основании полученных прогнозов;
- 3) генерирование деревьев решений на случайных подвыборках временных промежутков и вывод наиболее часто встречаемых в прогнозах деревьев акций в качестве окончательного результата; 4) проверка точности алгоритма.

Модель ARIMA. Интегрированная модель авторегрессии – скользящего среднего (ARIMA) – это модель авторегрессии – скользящего среднего (ARMA(p,q)), которая применяется не к значениям уровней ряда, а к разностям порядка d [7, 8] (для финансовых временных рядов достаточно $d = 1$). Модель ARMA (p, q) представляется следующим уравнением (5).

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q \beta_i \varepsilon_{t-i}, \quad (5)$$

где y_t – текущий уровень ряда; y_{t-i} – уровень ряда, отстающий на лаг i , ε_t – ошибка модели скользящего среднего; ε_{t-i} – ошибка скользящего среднего, отстающая на лаг $t-i$; α_i, β_i – параметры модели.

Первая сумма модели является компонентой авторегрессии, а вторая – компонентой скользящего среднего. Для модели ARIMA формула (4) преобразовывается следующим образом:

$$\Delta_t^d = \sum_{i=1}^p \alpha_i \Delta_{t-i}^d + \varepsilon_t + \sum_{i=1}^q \beta_i \varepsilon_{t-i}, \quad (6)$$

где Δ^d – разность порядка d .

При построении моделей ARIMA и ARMA возникает проблема определения ошибки ε_t (шумовой компоненты), решение которой предложено нами в работах [9, 10]. Для прогнозирования на фондовых рынках достаточно модели ARIMA порядков (1, 1, 1).

Дерево решений. Алгоритм «дерево решений» применяется в машинном обучении для задач классификации и регрессии [11]. Задача классификации состоит в том, чтобы разнести итоговый признак по двум или более категориям. К таким задачам относятся, к примеру, решение о выдаче кредита, выбор одаренных студентов и т.д. Сформулированная нами задача выбора акций для инвестиционного портфеля также относится к задаче классификации. На стадии обучения дерева итоговый признак уже разнесен по категориям. В корневой узел дерева выбирается один из факторных признаков, затем следует разветвление дерева, в зависимости от значений признака. Далее в следующих узлах выбираются другие факторные признаки. Процесс повторяется, пока дерево не достигнет заданной глубины. Последний уровень дерева дает окончательное разделение по категориям, и его узлы называются листья. Каждый узел содержит информацию о количестве элементов выборки в нем, а также метрику ошибки. Для классификации такими метриками выступают энтропия и коэффициент Джини [12], а для регрессии показатели отклонения предсказанных значений от фактических, такие как средняя квадратичная ошибка, средняя абсолютная ошибка и др. Как правило, глубина дерева выбирается такой, при которой ошибка перестает

уменьшаться. Для задачи классификации в узлах дерева решений могут находиться как категориальные, так и количественные признаки.

После обучения дерево решений проверяется на тестовой выборке. Фактические значения итогового признака не подаются алгоритму, а сравниваются со значениями, предсказанными деревом. На данном этапе для дерева классификации считается матрица ошибок, которая показана табл. 1.

Таблица 1. Матрица ошибок
Table 1. Confusion Matrix

Predictions	True Labels	
	Yes	No
	Yes	True Positive Labels
	No	False Negative Labels

На побочной диагонали матрицы ошибок находятся верно предсказанные отклики, на главной – неверно. По квадрантам матрицы ошибок рассчитываются метрики качества модели классификации. Наиболее распространены следующие метрики:

- **Precision**. Показывает долю правильно предсказанных положительных ответов в общем количестве предсказанных положительных ответов.
- **Recall**. Показывает долю положительных ответов, предсказанных алгоритмом, в общем количестве фактических положительных ответов.
- **Accuracy**. Показывает долю точных прогнозов по обоим классам. Алгоритм классификации считается приемлемым, если Accuracy составляет более 50%, иначе алгоритм классифицирует не лучше случайного выбора.
- **True Positive Rate**. Совпадает с метрикой Recall.
- **True Negative Rate**. Показывает долю верно предсказанных отрицательных ответов в общем количестве отрицательных ответов алгоритма.

Расчет перечисленных метрик представлен формулами 7–10.

$$Precision = \frac{TP}{TP+FP}, \quad (7)$$

$$Recall = \frac{TP}{TP+FN}, \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}, \quad (9)$$

$$TNR = \frac{TN}{TN+FN}, \quad (10)$$

где TP – количество откликов, которые алгоритм отнес к положительным, и при этом они действительно положительные, FP – количество откликов, которые алгоритм отнес к положительным, и при этом они отрицательные, TN – количество откликов, которые алгоритм отнес к отрицательным, и при этом они действительно отрицательные, FN – количество откликов, которые алгоритм отнес к отрицательным, и при этом они положительные [13].

Отметим, что важность критериев зависит от поставленной задачи. Например, при выборе студентов для стипендиальной программы важнее TNR , а для нашей задачи инвестиционного портфеля важнее метрики $Precision$ и $Recall$ (положительные отклики важнее отрицательных). Нам важно с одной стороны, чтобы в нашем портфеле среди отобранных акций было как можно больше действительно выросших в цене, что показывает метрика $Precision$, с другой стороны, чтобы растущие акции не были нами упущены, что показывает метрика $Recall$. Что касается метрики Accuracy, то она общая и ее необходимо считать при любом роде задаче классификации. Таким образом, для оценки качества модели мы выбираем метрики $Precision$, $Recall$ и $Accuracy$.

Пример дерева решений, построенного с помощью встроенных функций языка программирования R, показан на рис.1. На узлах дерева показано количество положительных и отрицательных ответов, а также количество элементов, соответствующих условиям,

которые показаны на ветвях дерева. Алгоритм обучается на фактических данных и затем сам принимает решения на тестовой выборке.

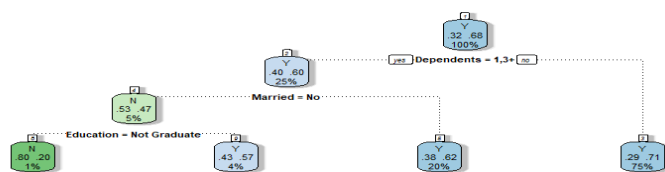


Рис. 1 - Пример дерева решений

Fig. 1 - Decision Tree example

Само по себе дерево решений является слабым алгоритмом классификации и прогнозирования. Для улучшения качества используются ансамблевые методы [14], основанные на построении множества деревьев решений. К таким алгоритмам относятся случайный лес [15] и градиентный бустинг [15].

Случайный лес. Основная идея алгоритма «случайный лес» – это построение множества решающих слабых алгоритмов – решающих деревьев и выбор в случае классификации тех объектов, за которые проголосует наибольшее количество деревьев, а в случае регрессии – среднего значения по всем деревьям. Алгоритм был разработан в 2001 году Л. Брейманом и А. Картер [17]. В основе алгоритма заложен ансамблевый метод бэггинг [18] – построение одинаковых моделей, результаты которых усредняются (в случае классификации – выбираются результаты, получившие большинство голосов).

Случайные подвыборки из исходного множества признаков и данных генерируются методом бутстреп – случайного генерирования из исходной выборки множества псевдовыборок, в которых обязательно присутствуют значения из исходной выборки, но при этом они могут повторяться. Выбор осуществляется с возвращением. В алгоритме случайного леса бутстреп используется на самом первом шаге: из исходной выборки генерируются псевдовыборки, и по каждой из них строится дерево решений. Признаки, по которым расщепляются значения признака, также в каждом дереве выбираются случайным образом. Как правило, при реализации алгоритма задается количество деревьев, количество признаков для обучения, скорость обучения и глубина дерева. Существуют и другие параметры для более тонкой настройки, например, минимальное количество признаков, по которым происходит расщепление. Критерии оценки точности алгоритма выбираются такие же, как и для одиночных деревьев решений. К примеру, для нашей задачи классификации — это метрики Precision, Recall и Accuracy [19].

Метод «случайный лес» используется в настоящее время без изменений и усовершенствований и до сих пор успешно применяется в задачах классификации и регрессии [20]. «Случайный лес» показывает сильные результаты во всех задачах машинного обучения, кроме распознавания изображений. В настоящее время набирает популярность альтернативный алгоритм градиентного бустинга – последовательное построение решающих деревьев, каждое последующее из которых дает лучшие результаты, чем предыдущее, однако алгоритм случайного леса продолжает активно использоваться. На наш взгляд, его преимущество перед градиентным бустингом заключается в простоте самостоятельной разработки без использования встроенных в программные пакеты библиотек, что дает возможность модифицировать алгоритм под конкретные задачи.

Обсуждение результатов. Модифицированный случайный лес. Алгоритм разработан авторами на языке программирования Python. Для расчета метрик точности (Precision, Recall и Accuracy) было осуществлено разделение исходной выборки на тренировочную и тестовую. За тренировочный период было взято 80% от исходной выборки, за тестовый – 20%. Прогноз осуществлялся на момент окончания тестового периода.

В первую очередь была создана функция, осуществляющая прогноз методом ARIMA, подробно представленная и протестированная в наших предыдущих работах [9].

Затем строилась функция, осуществляющая прогноз наивным методом (обычное среднее арифметическое тренировочной выборки). Пример прогноза наивным методом и методом ARIMA для акции Газпром за период с февраля 2023 г. по февраль 2024 г. представлен в соответствии с рис. 2, 3.

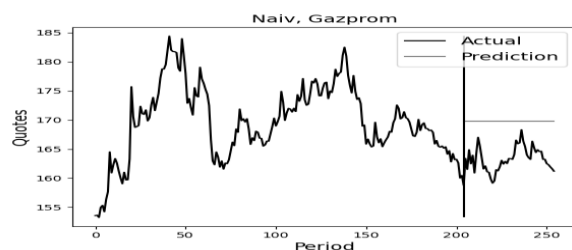


Рис. 2 - Пример прогноза на основе применения наивного метода
 Fig. 2 - Naïve method based forecast example

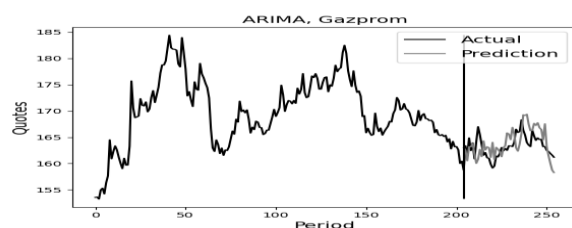


Рис. 3 - Пример прогноза на основе применения метода ARIMA
 Fig. 3 - ARIMA method based forecast example

Метод ARIMA как правило показывает лучшие результаты. Так, в соответствии с проиллюстрированным примером метод ARIMA показал среднюю ошибку в 0,5%, тогда как наивный прогноз – 4%. Однако, для разветвления дерева нам необходимо как минимум два метода. Более сложные, чем наивный прогноз, методы прогнозирования, а также включение большего количества методов, существенно увеличат время исполнения алгоритма, поэтому мы остановились на двух выше представленных методах.

Дерево решений было построено следующим образом: в корне дерева осуществлялся прогноз наивным методом (среднее арифметическое). Если прогноз показывал падение цены акции, то акция исключалась из портфеля. В противном случае акция переходила на следующий уровень дерева, на котором осуществлялся прогноз методом ARIMA. Если прогноз показывал рост, акция оставалась в портфеле, в противном случае – исключалась. Структура дерева показана на рис. 4.



Рис. 4 - Разработанное дерево решений
 Fig. 4 - Developed decision tree

Функция дерева на выходе возвращала два списка акций: принятых и не принятых в портфель. Для принятия решения, какой из двух методов разместить в корень дерева использовался коэффициент Джини, показывающий степень неопределенности, который рассчитывался по каждому из двух возможных разбиений. Выбирался метод с меньшим значением коэффициента. Далее необходимо составить из подобных деревьев «случайный лес». «Лес» должен соответствовать двум условиям.

Во-первых, поскольку мы имеем всего два фактора, по которым строится дерево, выбирать случайным образом факторы мы не можем, и следует учесть оба фактора на каждом дереве. Во-вторых, случайные подвыборки должны состоять из следующих друг

за другом элементов начальной выборки, так как мы строили прогноз. Для генерации случайных подвыборок с учетом указанных условий мы использовали два случайных числа: индекс исходной выборки k , начиная с которого мы выбираем подвыборку, и длина подвыборки l . Очевидно, что длина подвыборки не должна превышать разницу между длиной исходной выборки и k . Также мы задали значение l не меньше 50.

Чтобы определить количество подвыборок и, соответственно, количество деревьев, мы рассчитывали метрику Accuracy на этапе обучения. Оптимальным было выбрано то количество деревьев, при достижении которого метрика Accuracy перестала меняться. Таким образом, было сгенерировано 300 подвыборок. Далее подсчитывались «голоса» деревьев по каждой акции, входящей в индекс и, если количество «голосов» оказывалось меньше медианного значения, акция исключалась из портфеля, в противном случае – принималась в портфель.

Все вышеперечисленные действия производились на тренировочной выборке. Далее, для того, чтобы проверить точность алгоритма необходимо сравнить последнее значение цены отобранных акций на тестовой выборке с последним значением на тренировочной выборке. Если первое превышает второе, то мы ставили метку TRUE, в противном случае – метку FALSE. После проставления меток по всем акциям индекса рассчитывались метрики Precision, Recall и Accuracy.

Разработанный нами алгоритм был протестирован на акциях, входящих в индексы Мосбиржи (47 инструментов) и NASDAQ (231 инструмент). Для получения котировок с бирж и взаимодействия API Московской биржи и Yahoo Finance нами были разработаны специальные функции.

Отметим, что важно было протестировать алгоритм на разных временных периодах и временных промежутках, так как при тестировании на одном-двух периодах была возможность просто попасть в рост рынка. По этой причине мы взяли три временных промежутка: 2021-2022, 2022-2023 и 2023-2024 гг., с интервалами 05 года и 1 год. Результаты применения алгоритма «Модифицированный случайный лес» в виде метрик Precision, Recall и Accuracy представлены в соответствии с табл. 2.

Таблица 2. Результаты применения алгоритма «Модифицированный случайный лес»
Table 2. Results of applying the algorithm "Modified Random Forest"

Показатель	Период					
	16.02.23 – 16.02.24	01.03.22 – 01.03.23	26.12.21 – 26.12.22	16.06.23 – 16.02.24	01.07.22 – 01.03.23	26.06.21 – – 26.12.22
Precision, MOEX, %	77	64	86	67	76	100
Precision, NASDAQ, %	64	82	96	90	82	50
Recall, MOEX, %	61	58	59	62	53	59
Recall, NASDAQ, %	95	55	55	98	51	61
Accuracy, MOEX, %	61	56	62	59	54	62
Accuracy, NASDAQ, %	51	52	58	65	51	59

Для анализа были выбраны два индекса – российский индекс Мосбиржи из малого количества инструментов и американский NASDAQ с существенно большим количеством акций. В соответствии с результатами, представленными в таблице 1, в большинстве случаев алгоритм показал хорошие результаты – он выбирал для портфеля как минимум 64% прибыльных акций. Метрика Accuracy во всех случаях составила более 50%, что говорит о том, что алгоритмом имеет смысл пользоваться, поскольку он работает лучше, чем случайный выбор. Мы выбрали разные временные промежутки, чтобы показать работу алгоритма при разных состояниях рынках. Например, 2022 год был годом падения российского рынка. При этом результаты по отечественному индексу в этот период достаточно хороши.

Сравним с представленным в работе методом «Модифицированный случайный лес» классический метод формирования инвестиционного портфеля – решение задачи

Марковитца. Для решения данной задачи также был разработан алгоритм на языке программирования Python. Оптимальное решение задачи формировалось по формуле 2.

Отметим, что решение задачи мы представили не в классическом виде – нам нужны бинарные отклики (брать / не брать акцию в портфель). Таким образом, мы не считали доли отобранных акций, а только проверяли включенные в портфель Марковитца акции по метрикам Precision, Recall и Accuracy.

В табл. 3 показаны аналогичные представленным в табл. 2 показатели, полученные при применении задачи Марковитца. Согласно информации, представленной в табл. 2, результаты классической модели не лучше, а во многих случаях и хуже результатов разработанной нами модели. Метрика Ассурасу также, как и по модели случайного леса показывает результат свыше 50% во всех случаях.

Таблица 3. Результаты применения алгоритма «Задача Марковитца»

Table 3. Results of applying the algorithm "Markowitz problem"

Показатель	Период					
	16.02.23 – 16.02.24	01.03.22 – 01.03.23	26.12.21 – 26.12.22	16.06.23 – 16.02.24	01.07.22 – 01.03.23	26.06.21 – 26.12.22
Precision, MOEX, %	79	85	86	66	84	65
Precision, NASDAQ, %	49	90	78	85	83	32
Recall, MOEX, %	53	65	60	49	60	71
Recall, NASDAQ, %	51	51	49	49	53	55
Accuracy, MOEX, %	46	60	61	46	57	66
Accuracy, NASDAQ, %	51	52	49	50	51	54

Однако метрика Recall по американскому индексу за все периоды достаточно слабая, а метрика Precision по нему же принимает низкие значения в промежутке с 16.02.23 по 16.02.24 и 26.06.21 по 26.12.22. Из разницы в метриках на основе применения указанных методов следует еще один интересный вывод: разные методы выбирают разные акции в портфель, независимо от того, в каком направлении движется рынок.

Так, в самый кризисный из представленных для нашей страны период с 2021 по 2022 гг. метод «Модифицированный случайный лес» собрал прибыльный портфель. Отсюда следует, что даже при падении рынка, можно найти работающие методы, каковым является представленный нами алгоритм «Модифицированный случайный лес».

Вывод. В работе были рассмотрены классические подходы к формированию портфеля финансовых инструментов в контексте их достоинств и недостатков. Основной идеей исследования выступил подход к формированию портфеля, основанный на современном методе машинного обучения – «случайный лес». В качестве факторов, определяющих деревья решений, нами были взяты наивный прогноз по среднему арифметическому и прогноз методом ARIMA.

Данный алгоритм, как и известный «метод рюкзака», в отличие от классических методов, не определяет веса акций в портфеле, а определяет, включать ли акцию в портфель. Алгоритм был реализован на языке программирования Python и протестирован на двух индексах – российском индексе Мосбиржи и американском индексе NASDAQ – на трех временных промежутках за периоды один год и полгода. В качестве метрик выбрали наиболее распространенные метрики задачи бинарной классификации – Precision, Recall и Accuracy. На большей части выбранных временных промежутков алгоритм показал хорошие результаты. Метрика Ассурасу во всех случаях показала значение более 50%, из чего следует применимость алгоритма к поставленным задачам.

Для сравнения осуществили аналогичный отбор акций, используя классический алгоритм оптимизационной задачи Марковитца. Он также показал неплохие результаты, но на большинстве временных промежутков разработанный нами алгоритм показал лучшие результаты. Код алгоритма разработан и зарегистрирован в формате свидетельства о государственной регистрации программы для ЭВМ «Программа отбора акций для

инвестиционного портфеля акций индекса Мосбиржи на основе модифицированного алгоритма «Случайный лес».

Библиографический список:

1. Боди, З., Кейн А., Маркус А. Принципы инвестиций: пер. с англ. Москва, Издательский дом «Вильямс», 2005. – 984 с.. ISBN 9785845913111
2. Касимов Ю.Ф. Финансовые инвестиции. – Москва, Анкил. – 2008. – 230 с.
3. Иванченко И.С., Осей Д.Д. Оптимизация структуры российских золотовалютных резервов при помощи модели Блэка — Литтермана // Финансовый журнал. – 2018. – № 1 (41). DOI 10.31107/2075-1990-2018-1-26-38
4. Шарп У., Бэйли А. Инвестиции Пер. с англ. – Москва, ИНФРА-М, 2009 – 1027 с. ISBN 9785160167893
5. Косоруков О.А., Мищенко А.В. Исследование операций. – Москва: Экзамен. – 2003. – 448 с.
6. Зиненко А.В., Мартынова Т.А. Модифицированный алгоритм метода рюкзака для решения обратной задачи Марковитца// Финансовая экономика, 2024. – № 6. – С. 136 – 139.
7. Бокс Дж., Дженкинс Г. Анализ временных рядов: прогноз и управление. – Москва, Мир. – 1974 – 820 с. ISBN 2100013644333
8. Вьюгин В.В. Математические основы машинного обучения и прогнозирования. – Москва, МЦНМО. – 2024. – 484 с. ISBN 9785443920146
9. Zinenko A., Stupina A. Financial time series forecasting methods// ITM Web Conf. II International Workshop “Hybrid Methods of Modeling and Optimization in Complex Systems” (HMMOCS-II 2023). – 2024. – Vol 59. 02005. DOI 10.1051/itmconf/20245902005
10. Chou J.-Sh., Chen K.-E. Optimizing investment portfolios with a sequential ensemble of decision tree-based models and the FBI algorithm for efficient financial analysis// Applied Soft Computing. – 2024. – Vol.158. – 111550. DOI 10.1016/j.asoc.2024.111550
11. Марченко А.Л. Python: большая книга примеров. – Москва, Издательство Московского университета. – 2023. – 361 с.
12. Бринк Х., Ричардс Д., Феверлоф М. Машинное обучение. – Санкт-Петербург, Питер Айлиб. – 2017. – 336 с. ISBN 9785496029896
13. Рашка С., Мирджалили Р. Python и машинное обучение. – Москва, Packt. – 2020. – 656 с. ISBN 9785970604090
14. Картер Д. Искусственный интеллект. Машинное обучение. – Москва, «Автор». – 2024. – 268 с.
15. Habbab F., Kampouridis M. Anin-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs// Systems With Applications. – 2024. – Vol.235. –121102. DOI 10.1016/j.eswa.2023.121102
16. Jun S.Y. et al. Stock investment strategy combining earnings power index and machine learning// International Journal of Accounting Information Systems. – 2022. –Vol. 47. –100576. DOI 10.1016/j.acinf.2022.100576
17. Behera J. et al. Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets// Engineering Applications of Artificial Intelligence. – 2023. – Vol. 120. – 105843. DOI 10.1016/j.engappai.2023.105843
18. 190. Parisi L., Manaog M.L. Optimal Machine Learning- and Deep Learning-driven algorithms for predicting the future value of investments: A systematic review and meta-analysis// Engineering Applications of Artificial Intelligence. – 2025. – Vol. 142. – 109924. DOI 10.1016/j.engappai.2024.109924
19. Ramirez H.E., Serrano R. Optimal investment with insurable background risk and nonlinear portfolio allocation frictions// Applied Mathematics and Computation. – 2025. – Vol. 485 – 129023. DOI 10.1016/j.amc.2024.129023.
20. Lin W., Taamouty A. Portfolio selection under non-gaussianity and systemic risk: A machine learning based forecasting approach// International Journal of Forecasting. – 2024. – Vol. 40 – Pp. 1179 – 1188. DOI 10.1016/j.ijforecast.2023.10.007.

References:

1. Bodie, Z., Kane, A., Marcus, A. Investment Principles: trans. from English. Moscow: Williams Publishing House. 2005. 984 p. ISBN 9785845913111
2. Kasimov, Yu.F. Financial Investments. Moscow: Ankil. 2008. 230 p.
3. Ivanchenko, I.S., Osei, D.D. Optimizing the Structure of Russian Gold and Foreign Exchange Reserves Using the Black-Litterman Model. *Financial Journal*. 2018; 1 (41): 26–38. DOI 10.31107/2075-1990-2018-1-26-38 (In Russ)
4. Sharpe, W., Bailey, A. Investments. Trans. from English. Moscow: INFRA-M. 2009. 1027 p. ISBN 9785160167893 (In Russ)
5. Kosorukov O.A., Mishchenko A.V. Operations Research. Moscow: Examen. 2003. 448 p.

6. Zinenko A.V., Martynova T.A. Modified knapsack method algorithm for solving the inverse Markowitz problem. *Financial Economics*. 2024; 6: 136 –139 (In Russ)
7. Box J., Jenkins G. *Time Series Analysis: Forecast and Management*. Moscow: Mir. 1974. 820 p. ISBN 2100013644333 (In Russ)
8. Vyugin V.V. *Mathematical Foundations of Machine Learning and Forecasting*. Moscow: MCNO. 2024. 484 p. ISBN 9785443920146 (In Russ)
9. Zinenko A., Stupina A. Financial time series forecasting methods. *ITM Web Conf. II International Workshop "Hybrid Methods of Modeling and Optimization in Complex Systems" (HMMOCS-II 2023)*. 2024; 59: 02005. DOI 10.1051/itmconf/20245902005
10. Chou J.-Sh., Chen K.-E. Optimizing investment portfolios with a sequential ensemble of decision tree-based models and the FBI algorithm for efficient financial analysis. *Applied Soft Computing*. 2024; 158: 111550. DOI 10.1016/j.asoc.2024.111550
11. Marchenko A.L. *Python: a big book of examples*. Moscow: Moscow University Press. 2023;361 (In Russ)
12. Brink H., Richards D., Feverloff M. *Machine Learning*. St. Petersburg: Peter Ilib. 2017. 336 p. ISBN 9785496029896 (In Russ)
13. Rashka S., Mirjalili R. *Python and Machine Learning*. Moscow: Packt. 2020. 656 p. . ISBN 9785970604090 (In Russ)
14. Carter D. *Artificial Intelligence. Machine Learning*. Moscow: "Author". 2024. 268 p. (In Russ)
15. Habbab F., Kampouridis M. Anin-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. *Systems With Applications*. 2024; 235:121102. DOI 10.1016/j.eswa.2023.121102
16. Jun S.Y. et al. Stock investment strategy combining earnings power index and machine learning. *International Journal of Accounting Information Systems*. 2022; 47: 100576. DOI 10.1016/j.accinf.2022.100576
17. Behera J. et al. Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*. 2023; 120:105843. DOI 10.1016/j.engappai.2023.105843
18. 190. Parisi L., Manaog M.L. Optimal Machine Learning- and Deep Learning-driven algorithms for predicting the future value of investments: *A systematic review and meta-analysis// Engineering Applications of Artificial Intelligence*. 2025; 142: 109924. DOI 10.1016/j.engappai.2024.109924
19. Ramirez H.E., Serrano R. Optimal investment with insurable background risk and nonlinear portfolio allocation frictions. *Applied Mathematics and Computation*. 2025; 485: 129023. DOI 10.1016/j.amc.2024.129023
20. Lin W., Taamouty A. Portfolio selection under non-gaussianity and systemic risk: A machine learning based forecasting approach. *International Journal of Forecasting*. 2024; 40: 1179 – 1188. DOI 10.1016/j.ijforecast.2023.10.007

Сведения об авторе:

Анна Викторовна Зиненко, кандидат технических наук, доцент, кафедра «Экономическая и финансовая безопасность», anna-z@mail.ru. ORCID 0000-0002-4212-0321

Information about authors:

Anna V. Zinenko, Cand. Sci. (Eng.), Assoc. Prof.; Department of Economic and Financial Security, anna-z@mail.ru. ORCID 0000-0002-4212-0321

Конфликт интересов/Conflict of interest.

Автор заявляет об отсутствии конфликта интересов/The author declare no conflict of interest.

Поступила в редакцию/Received 24.03.2025.

Одобрена после рецензирования/Revised 25.04.2025.

Принята в печать/Accepted for publication 20.07.2025.