

УДК 004.85

DOI: 10.18413/2518-1092-2025-10-3-0-2

Назаров Т.Р.

ФОРМИРОВАНИЕ ИНВЕСТИЦИОННЫХ ПОРТФЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ КЛАСТЕРНОГО АНАЛИЗА

Российский экономический университет имени Г.В. Плеханова,
Стремянный переулок, 36, г. Москва, 115054, Россия

e-mail: TimNazarovya@gmail.com

Аннотация

Исследование посвящено проверке гипотезы о возможности создать диверсифицированный портфель акций путем их кластеризации по показателям риска и устойчивости. В рамках данной проблемы использовались три метода кластеризации: метод К-средних, модель гауссовой смеси и иерархическая кластеризация методом Уорда. Также было отобрано три признака, по которым проводилась кластеризация акций: индекс Бета, волатильность и ROE. По данным с 2018 по 2023 год за каждый период была проведена кластеризация с подбором оптимального количества кластеров, и были созданы портфели финансовых активов по заранее определенным правилам. В качестве альтернативы за каждый год были созданы портфели акций без предварительной кластеризации по правилам портфельной теории Марковица. После этого каждый портфель был протестирован в течение следующего года. Результаты на основе метрик доходности и риска показывают, что гипотеза имеет право на существование. Портфели, созданные с предварительной кластеризацией, превосходили по доходности классические варианты. Проблема имеет простор для дальнейшего исследования. В частности, иной набор признаков для сегментации акций или применение методов кластеризации в сочетании с другими алгоритмами машинного обучения или нейронных сетей может также дать качественные результаты.

Ключевые слова: кластеризация; акции; волатильность; Индекс Бета; К-средних; иерархическая кластеризация; метод гауссовых смесей; портфельная теория Марковица

Для цитирования: Назаров Т.Р. Формирование инвестиционных портфелей с использованием кластерного анализа // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 20-30. DOI: 10.18413/2518-1092-2025-10-3-0-2

Nazarov T.R.

FORMATION OF INVESTMENT PORTFOLIOS USING CLUSTER ANALYSIS

Plekhanov Russian University of Economics,
36 Stremyanny Lane, Moscow, 115054, Russia

e-mail: TimNazarovya@gmail.com

Abstract

The study is devoted to testing the hypothesis about the possibility of creating a diversified portfolio of stocks by clustering them according to risk and sustainability indicators. Within the framework of this problem, three clustering methods were used: the K-means method, the Gaussian mixture model, and hierarchical clustering by the Ward method. We also selected three criteria for stock clustering: Beta index, volatility, and ROE. According to the data from 2018 to 2023, clusterization was carried out for each period with the selection of the optimal number of clusters, and portfolios of financial assets were created according to predefined rules. Alternatively, stock portfolios were created for each year without prior clustering according to the rules of Markowitz portfolio theory. After that, each portfolio was tested over the next year. The results based on the metrics of profitability and risk show that the hypothesis has a right to exist. Portfolios created with pre-clustering outperformed the classic options in terms of profitability. The problem has scope for further investigation. In particular, a different set of features for stock segmentation or the use of

clustering methods in combination with other machine learning or neural network algorithms can also produce high-quality results.

Keywords: clustering; stocks; volatility; Beta index; K-means; hierarchical clustering; Gaussian mixture method; Markowitz portfolio theory

For citation: Nazarov T.R. Formation of Investment Portfolios Using Cluster Analysis // Research result. Information technologies. – Т.10, №3, 2025. – P. 20-30. DOI: 10.18413/2518-1092-2025-10-3-0-2

ВВЕДЕНИЕ

На конец 4 квартала 2024 года, более трети экономически активного населения имеет брокерские счета и по итогам года клиентская база брокеров выросла на 36% за счет новичков [1]. С помощью рынка ценных бумаг можно получать хорошую прибыль, но для этого необходимо эффективно управлять инвестиционным портфелем, анализировать динамику фондового рынка, рассчитывать уровень доходности и просчитывать возможные риски. Из-за непредсказуемой ситуации на рынке, а также предубеждений, часто наблюдаемых в поведении людей, связанных с инвестированием, и нелогичности принятия решений, создание успешных портфелей финансовых активов и управление ими является сложной задачей. Достижение высокой доходности при низком риске – идеальная, но кажущаяся невозможной задача. Этот процесс называется процессом диверсификации. Современная портфельная теория утверждает, что диверсификация активов является наиболее эффективным способом получения портфеля активов с низкими показателями риска и относительно высокими показателями прибыли. На данный момент было представлено множество методов создания портфеля акций, из которых наиболее известным является метод Марковица, который был предложен в 1951 году [2].

Кластеризация – один из наиболее важных методов интеллектуального анализа данных для извлечения полезной информации из массивов данных большой размерности. Кластеризация акций может помочь при построении диверсифицированного портфеля ценных бумаг. Например, путем распределения капитала между различными кластерами акций инвестор может снизить общий уровень риска портфеля. Если один сектор или регион испытывает трудности, другие кластеры могут остаться стабильными или даже вырасти, что помогает сгладить потери.

Данная проблема не нова, и ей посвящено много исследований. Исследования подтверждают, что с помощью различных методов кластеризации возможно получить устойчивые группы акций [3]. Портфели, созданные с помощью сравнения временных рядов [12, 13] и анализа динамических показателей [4, 11] акций, демонстрировали высокую доходность и умеренный риск. Однако, в упомянутых исследованиях портфели не сравнивались с оптимизированными портфелями, созданными без предварительной кластеризации акций. Данное исследование будет рассматривать в качестве показателей как динамические показатели акций, так и финансовые показатели эмитентов акций, а показатели доходности и риска портфелей будет сравниваться с классическими вариантами.

Целью данной работы является проверка гипотезы о том, что кластеризация акций по определенным показателям риска и доходности может помочь создать диверсифицированный портфель акций, который может превзойти классические альтернативы по доходности, при этом представляя меньший риск для инвестора.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

За основу брались акции, составляющие в период с 2018 по 2023 год индекс широкого рынка на Московской Бирже. Индекс отражает общее состояние динамики рынка акций. Фондовые индексы имеют весомую роль при оценке конъюнктуры фондового рынка, позволяют анализировать его состояние в прошлые периоды времени, выявлять тенденции развития, а также делать прогнозы на будущее. Индекс акций широкого рынка Московской Биржи включает в себя 100 ценных бумаг, отобранных на основании критериев ликвидности, капитализации и доли,

находящейся в свободном обращении, и является основой для формирования баз расчета остальных индексов Московской Биржи [5]. С учетом полноты доступных данных, распределение количества акций компаний, доступных для добавления в портфель по годам, представлено в Таблице 1:

Таблица 1

Количество доступных акций в рамках каждого года

Table 1

Number of shares available within each year

Год	Количество акций
2018	81
2019	85
2020	78
2021	71
2022	73
2023	78

Сравнение портфелей будет происходить исключительно в рамках каждого года, поэтому различное количество доступных акций от года к году не является критичным для проведения исследования. Для проведения кластерного анализа акций были выбраны следующие показатели:

– мультипликатор ROE (return on equity). Данный мультипликатор измеряет прибыльность компании путем сравнения ее чистой прибыли с собственным капиталом. Показывает, насколько эффективно компания генерирует прибыль за счет своего собственного капитала. Более высокий показатель рентабельности инвестиций, как правило, желателен, поскольку это говорит о том, что компания более эффективно использует свой собственный капитал для получения прибыли.

– Коэффициент Бета. Отражает рискованность вложения по отношению к рынку и показывает чувствительность изменения доходности акции к изменению доходности рынка [6]. Рассчитывается по следующей формуле:

$$\beta_a = \frac{cov(r_a, r_p)}{var(r_p)}, \quad (1)$$

где r_a – доходность акции, а r_p – доходность рынка.

В качестве актива, по которому оценивается доходность рынка, рассматривался индекс Московской Биржи;

– волатильность в процентах. Волатильность, в контексте акций, рассчитывается как стандартное отклонение изменения цены за период времени. Акции с более высокой волатильностью, как правило, являются более рискованными инвестициями, поскольку их цены с течением времени колеблются более резко.

Эти признаки вместе охватывают разные аспекты риска и доходности, что делает их хорошей базой для кластеризации: фундаментальные свойства (ROE), рыночная чувствительность (Beta), внутренняя нестабильность (волатильность). Аналогичные переменные традиционно используются в работах по факторному инвестированию [15,16], а также при построении многомерных моделей риска. Пример наблюдений из набора данных представлен на Рисунке 1:

	Год	Ticker	ROE, %	Beta	Volatility
0	2018	AFKS	6.5	0.296480	1.635736
1	2018	AFLT	-9.0	0.163191	2.001954
2	2018	AGRO	12.6	0.288014	1.217860
3	2018	AKRN	22.2	0.409532	0.897271
4	2018	ALRS	36.5	0.317364	1.698126

Рис. 1. Пример наблюдений из набора данных
Fig. 1. Example of observations from the dataset

Колонка Ticker содержит краткое название котируемого на Московской Бирже инструмента. Для отображения многомерных данных в двумерном пространстве используется метод многомерного шкалирования. В общем случае метод позволяет таким образом расположить объекты в пространстве некоторой небольшой размерности, чтобы достаточно адекватно воспроизвести наблюдаемые расстояния между ними. Сходство между объектами обычно представляется в виде матрицы расстояний. Эта матрица показывает расстояния между каждой парой объектов в наборе данных. Отображение данных по годам представлено на Рисунке 2:

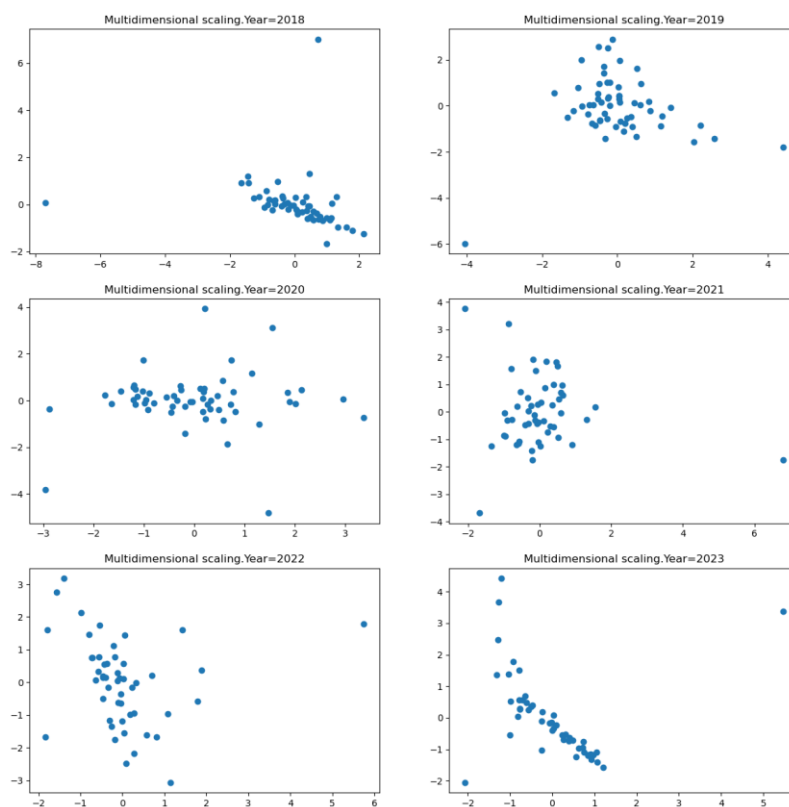


Рис. 2. Представление данных по годам
Fig. 2. Data presentation by year

Портфельные инвестиции – это вложение денежных средств в акции, облигации и другие ценные бумаги, выпущенные на рынок различными инвесторами (государством, корпорациями и частными компаниями), а портфель ценных бумаг – набор ценных бумаг, обеспечивающий для инвестора удовлетворительные количественно-качественные характеристики финансовых инструментов. Уменьшение риска вложений, как правило, достигается за счет диверсификации

(разделения) портфеля, то есть невысокие доходы по одним ценным бумагам компенсируются высокодоходными. На практике используют множество методик формирования оптимальной структуры портфеля ценных бумаг. Большинство из них основано на методике Марковица. Он впервые предложил математическую формализацию задачи нахождения оптимальной структуры портфеля ценных бумаг в 1951 году, за что позднее был удостоен Нобелевской по экономике.

Портфельная теория Марковица является многокритериальной оптимизационной задачей, в которой ищутся такие значения распределений долей активов в портфеле, которые одновременно обеспечивают максимально возможное значение ожидаемой доходности портфеля активов при минимально возможном значении риска портфеля активов (среднего стандартного отклонения) [7].

Ожидаемая доходность портфеля максимизируется следующим образом:

$$m_p = W^T * M \rightarrow \max. \quad (2)$$

Риск портфеля минимизируется следующим образом:

$$S_p^2 = W^T * V * W \rightarrow \min. \quad (3)$$

При ограничениях:

$$\sum_{i=1}^n W_i = 1 \quad (4)$$

$$w_i \geq 0 \forall i = 1 \dots n, \quad (5)$$

где $W_{[n \times 1]}$ – вектор искомых долей активов в портфеле, $M_{[n \times 1]}$ – вектор ожидаемых доходностей активов, отбираемых в портфель, $V_{[n \times n]}$ – матрица ковариаций доходности активов, а n – количество активов. Предполагается, что найденные в результате решения задачи, а именно веса активов, будут оптимальными и в следующих торгах. Поэтому, данная задача одновременно является задачей прогнозирования.

Кластеризация или кластерный анализ – это метод машинного обучения, который обрабатывает немаркированный набор данных и относится к методам обучения без учителя. Его цель – группировка объектов на подмножества или “кластеры” таким образом, чтобы объекты внутри каждого кластера были более тесно связаны друг с другом, чем объекты, отнесенные к другим кластерам. Формально, задачу кластеризации можно поставить следующим образом [14]: пусть X – совокупность наблюдений, Y – множество меток кластеров. На множестве X задана функция, описывающая расстояния между объектами $\rho(x, x')$. Также представлена обучающая выборка $X^m = \{x^1, \dots, x^m\}$. Необходимо разбить выборку на подмножества (кластеры), то есть каждому объекту $x_i \in X^m$ сопоставить метку $y_i \in Y$, таким образом, чтобы объекты внутри каждого кластера были близки относительно метрики ρ , а объекты из разных кластеров значительно отличались относительно метрики ρ . Таким образом, алгоритм кластеризации – функция $a: X \rightarrow Y$, которая любому объекту $x \in X$ ставит в соответствие идентификатор кластера $y \in Y$. Множество Y в некоторых случаях известно заранее, однако чаще ставится задача определить оптимальное число кластеров, с точки зрения того или иного критерия качества кластеризации.

Для кластеризации акций использовалось три различных метода кластеризации: метод k -средних, метод гауссовых смесей и метод иерархической кластеризации Уорда. Алгоритм K -средних является одним из самых популярных алгоритмов машинного обучения. Целью алгоритма является разбиение набора объектов на K кластеров таким образом, чтобы сумма квадратов расстояний между объектами и назначенным им средним значением по кластеру была сведена к минимуму. Каждому кластеру соответствует точка (центроид) в пространстве данных, которая является средним значением присвоенных ему наблюдений. Метод иерархической кластеризации Уорда, как следует из названия, представляет собой алгоритм, который строит иерархию кластеров. Работа алгоритмов иерархической кластеризации начинается с обработки каждого наблюдения как отдельного кластера. Затем последовательно выполняются следующие два

шага: идентифицируются два наиболее близких друг к другу кластера и объединяются два наиболее похожих кластера. Этот итеративный процесс продолжается до тех пор, пока все кластеры не будут объединены вместе или не будет достигнут критерий останова. В методе Уорда кластеры объединяются таким образом, чтобы минимизировать увеличение общей внутрикластерной дисперсии при объединении двух кластеров. Для определения оптимального значения порога расстояния, следует рассмотреть результаты, полученные при разных показателях и сделать выбор, основываясь на метриках качества кластеризации [8]. Метод Гауссовских смесей представляет собой алгоритм нечеткой кластеризации, где принадлежность каждого элемента тому или иному кластеру выражается в виде вероятности. Алгоритм предполагает, что точки данных внутри каждого кластера распределяются в соответствии с гауссовым распределением.

Для оценки схожести двух объектов использовалась евклидова метрика. Для двух точек в двумерном пространстве она вычисляется как квадратный корень из суммы разностей между соответствующими координатами двумя различными наблюдениями:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (6)$$

Евклидово расстояние хорошо проходит для работы с числовыми типами переменных.

Для оценки качества кластеризации и выбора оптимального количества кластеров использовался коэффициент силуэта. Коэффициент силуэта показывает, насколько правильно распределена каждая отдельная точка. Если коэффициент ближе к 0, то можно считать, что она находится на границе между двумя кластерами. Если он ближе к -1, то можно считать, что она неверно отнесена к определенному кластеру. Если коэффициент силуэта ближе к 1, то точка правильно назначена и может быть интерпретирована как принадлежащая к подходящему кластеру. Это очень интуитивно понятный способ оценки качества модели кластеризации. Данный показатель для всей выборки можно рассчитать следующим образом:

$$S = \frac{ncd - icd}{\max(icd, ncd)}, \quad (7)$$

где

icd – среднее внутрикластерное расстояние, то есть среднее расстояние между всеми членами кластера. Оно позволяет оценить, насколько плотно наблюдения расположены внутри каждого кластера.

ncd – среднее межкластерное расстояние, то есть среднее расстояние между каждым наблюдением в рассматриваемом кластере и наблюдениями в других кластерах.

Эмпирически установленное оптимальное количество кластеров в рамках каждого года, в зависимости от используемого метода кластеризации, представлено в Таблице 2:

Таблица 2

Оптимальное количество кластеров в зависимости от метода кластеризации и года

Table 2

Optimal number of clusters depending on clustering method and year

Метод	2018	2019	2020	2021	2022	2023
К-средних (K-means)	3	4	3	3	4	6
Метод Гауссовских смесей (GMM)	3	3	3	3	3	6
Иерархическая кластеризация Уорда (Ward)	4	4	5	4	4	6

Построение портфелей происходило по следующим правилам:

- вычислялось количество элементов, которое будет набрано из каждого кластера. Оно равнялось частному от деления нацело размера портфеля на количество кластеров. Если размер кластера меньше полученного числа, то из кластера брались все элементы;

- при применении алгоритма К-средних элементы из каждого кластера отбирались по убыванию расстояния от центроида своего кластера. Чем ближе элемент к центроиду – тем лучше он характеризует получившийся кластер. При применении метода Гауссовых смесей элементы из каждого кластера отбирались по убыванию вероятности принадлежности к своему кластеру. Чем выше вероятность – тем лучше элемент характеризует получившийся кластер. При применении метода иерархической кластеризации Уорда внутри каждого кластера находится медоид. Медоид – существующий элемент кластера, суммарное расстояние до которого от других элементов кластера минимально [9]. Элементы из кластера выбираются в портфель на основании расстояния до медоида. Чем меньше расстояние до медоида, тем лучше элемент характеризует получившийся кластер;

- остаток, если он возникает, добирался без привязки к кластеру, но по тому же правилу.

После отбора акций в портфели, внутри каждого портфеля для каждого актива назначался вес на основе портфельной теории Марковица с максимизацией индекса Шарпа и на основе данных о ежедневной динамике цен по каждому коду ценной бумаги, попавшей в портфель. Инвесторы часто используют коэффициент Шарпа для сравнения эффективности различных портфелей или инвестиционных стратегий. Он помогает инвесторам оценить, оправдывает ли дополнительная доходность, получаемая от портфеля, дополнительный риск, который они берут на себя. Вычисляется он по следующей формуле [10]:

$$SR = \frac{r_p - r_f}{\sigma_p}, \quad (8)$$

где r_p – средняя доходность портфеля, r_f – средняя безрисковая доходность, а σ_p – стандартное отклонение доходностей активов. Вес каждой акции в портфеле ограничивался диапазоном от 1 до 25%. Сумма, на которую формировался портфель – 6 миллионов 500 тысяч рублей. Акции покупались по цене на момент открытия биржи в первый торговый день следующего года. Зачастую вся сумма распределялась не полностью. В таких случаях остаток распределялся итеративно между акциями по убыванию веса.

ТЕСТИРОВАНИЕ ПОРТФЕЛЕЙ И АНАЛИЗ РЕЗУЛЬТАТОВ

Портфели сравнивались по годам. Так, портфели, созданные на основе данных за рассматриваемый год, в рамках эксперимента оценивались на основе данных, полученных за следующий год. Например, портфели, созданные на основе данных за 2018 год, оценивались по их показателям за 2019 год. В процессе оценки поведения портфеля не использовалась ребалансировка. Веса оставались неизменными, это дало возможность оценить качество диверсификации, полученной с помощью предварительной кластеризации.

Также было сформировано 5 портфелей без применения кластерного анализа. Они формировались следующим образом:

- в рамках каждого года у каждой акции брались данные по изменению цены за период с июня по декабрь. Всем акциям назначались веса, а потом отбирались 16 акций с наиболее высокими весами;

- отобранным бумагам назначались итоговые веса по тому же алгоритму в диапазоне от 0.01 до 0.25;

- акции приобретались по цене на момент открытия в первый торговый день следующего года на тот же бюджет.

Для сравнения портфелей были выбраны следующие метрики:

– доходность портфеля (в процентах) – отражает прирост или просадку портфеля в процентах по прошествии периода времени. Вычисляется по следующей формуле:

$$D = \frac{V_n - V_k}{V_k} * 100, \quad (9)$$

где V_n – начальный объем портфеля, V_k – конечный объем портфеля;

– отношение среднедневной доходности в течение года к годовой волатильности – данный коэффициент позволяет оценить эффективность портфеля с точки зрения соотношения дохода к риску. Более высокое значение может указывать на то, что портфель активов генерируют большую доходность при том же уровне риска:

$$RV = \frac{rp}{\sigma p}, \quad (10)$$

где rp – среднегодовая доходность портфеля, а σp – стандартное отклонение доходностей портфеля.

Результаты, полученные в ходе эксперимента, представлены в Таблице 3:

Таблица 3

Результаты тестирования портфелей ценных бумаг

Table 3

Results of securities portfolios testing

Метод	Год	Доходность, %	RV
K-means	2018	+16.5%	0.09
GMM	2018	+6.7%	0.04
Ward	2018	+12.6%	0.07
Classic	2018	+3.56%	0.03
K-means	2019	+53.37%	0.119
GMM	2019	+10.1%	0.03
Ward	2019	+24.3%	0.06
Classic	2019	+40.9%	0.105
K-means	2020	+61.8%	0.09
GMM	2020	+52.47%	0.15
Ward	2020	+80.11%	0.09
Classic	2020	+39.68%	0.12
K-means	2021	-10%	-0.01
GMM	2021	-37.6%	-0.05
Ward	2021	-21.5%	-0.02
Classic	2021	-10.76%	-0.007
K-means	2022	+29.1%	0.11
GMM	2022	+67.52%	0.15
Ward	2022	+28.14%	0.103
Classic	2022	+46%	0.09
K-means	2023	-7.84%	-0.02
GMM	2023	-14%	-0.03
Ward	2023	-9%	-0.018
Classic	2023	-16.24%	0.06

При создании портфелей на основе данных за 2018, 2019, 2020, 2022 год, портфели, созданные с помощью того или иного метода кластеризации, превзошли классический портфель как по общей

доходности, так и по отношению среднедневной доходности к годовой волатильности. Портфель, созданный на основе данных за 2021 год с помощью метода кластеризации К-средних, потерял в 2022 году чуть меньше, чем классический портфель, но, по сравнению с классическим портфелем, обладает худшим отношением среднегодовой доходности к годовой волатильности. Все портфели, созданные с помощью предварительной кластеризации на основе данных за 2023 год, в 2024 году потеряли меньше, чем классический портфель, но последний показал положительное отношение среднедневной доходности к годовой волатильности.

ЗАКЛЮЧЕНИЕ

Таким образом, в ходе исследования были собраны данные по акциям, составляющим индекс широкого рынка Мосбиржи в период с 2018 по 2023 год включительно. Данные включали в себя мультипликатор ROE, индекс Beta и годовую волатильность. В рамках каждого года была произведена кластеризация акций тремя различными методами и отобраны акции для составления трех портфелей. Акциям назначался вес согласно портфельной теории Марковица. Портфели сравнивались по общей доходности и отношению среднедневной доходности в течение года к годовой волатильности. В дальнейшем портфели были протестированы на последующем году. По полученным данным можно сделать вывод, что гипотеза о том, что кластеризация акций по определенным показателям риска и доходности (не ограничиваясь стандартными показателями доходности и волатильности), может помочь создать диверсифицированный портфель акций, имеет право на существование. Портфели, созданные на основе предварительной кластеризации тем или иным методом, превзошли классические альтернативы. По показателю отношения среднедневной доходности к годовой волатильности их превосходство проявилось не так наглядно. Также по итогам эксперимента нельзя явно определить наилучший метод кластеризации. При дальнейшем рассмотрении данного вопроса возможно как пересмотреть состав признаков, так и рассмотреть возможность применения других методов кластеризации или методов кластеризации в сочетании с другими алгоритмами машинного обучения или нейронных сетей.

Список литературы

1. Число физлиц с брокерскими счетами на МосБирже в 2024 году выросло на 5,4 млн // URL: <https://www.interfax.ru/business/1003027> (дата обращения: 26.05.2025).
2. Баюк Д., Баюк О., Берзин Д., Иванюк В. Практическое применение методов кластеризации, классификации и аппроксимации на основе нейронных сетей. – 1-е изд. – Москва: Прометей, 2020. – 448 с.
3. Дмитриев Н.Д. Кластеризация российских компаний технологического сектора на основе рыночных и финансовых показателей / Н.Д. Дмитриев, А.В. Тюлькова, В.В. Сорокожердьев // Вестник Челябинского государственного университета. – 2024. – № 10(492). – С. 116-127.
4. Тюхова Е.М. Использование кластерного анализа для формирования портфеля ценных бумаг в инвестиционных системах (робоедвайзерах) / Е.М. Тюхова, Д.С. Сизых // Управление развитием крупномасштабных систем (MLSD'2019): Материалы двенадцатой международной конференции, Москва, 01–03 октября 2019 года / Под общей редакцией С.Н. Васильева, А.Д. Цвиркуна. – Москва: Международный научно-исследовательский институт проблем управления РАН, 2019. – С. 210-220.
5. Индекс широкого рынка // URL: <https://www.moex.com/ru/index/MOEXBMI/about> (дата обращения: 24.05.2024).
6. Жданов И.Ю. Инвестиционная оценка проектов и бизнеса: учебное пособие / И.Ю. Жданов, В.Ю. Жданов. – Москва: Проспект, 2019. – 120 с.
7. Борисова Л.В., Сагаева И.Д. Модели оптимального инвестирования: учебно-методическое пособие, Саратов // URL: http://elibrary.sgu.ru/uch_lit/1347.pdf (дата обращения: 23.05.2025).
8. Кластеризация и частичное обучение // URL: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf> (дата обращения: 22.05.2025).
9. Тахир М.А., Ахмед А.У., Егорова Л.Д. Об алгоритмах для задачи К-медоид // Актуальные проблемы авиации и космонавтики. – 2021. – №1. – С. 194-196.

10. Деканова К.В., Гейт А.А. Методика формирования инвестиционного портфеля с учетом индивидуальных особенностей частного инвестора // Экономика: вчера, сегодня, завтра. 2023. – Том 13. – № 7А. – С. 311-329.

11. Nourahmadi M., Sadeqi H. Portfolio Diversification Based on Clustering Analysis // Iranian Journal of Accounting, Auditing and Finance. – 2023. – №7(3) – P. 1-16.

12. Szczepocki, Piotr Clustering companies listed on the Warsaw Stock Exchange according to time-varying beta // Econometrics. – 2023. – №23 – P. 63-79.

13. Lim T., Ng H. Financial Portfolio Management Based on Shaped-Based Unsupervised Machine Learning: A Dynamic Time Warping Baycenter Averaging Approach to International Markets and Periods of Downside Event Risks // The Journal of Investing. – 2023. – №32. – P. 74-96.

14. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. – 12th ed. – Springer, 2017. – 745 p.

15. Fama, E. F., & French, K. R. The Cross-Section of Expected Stock Returns // Journal of Finance. – 1992. – №47(2). – P. 427-465.

16. Asness C.S., Moskowitz T., Pedersen L. Value and Momentum Everywhere // Journal of Finance. – 2013. – 68. – P. 929-985.

References

1. The number of individuals with brokerage accounts on the Moscow Stock Exchange increased by 5.4 million in 2024. URL: <https://www.interfax.ru/business/1003027> (accessed: 05/26/2025).

2. Bayuk D., Bayuk O., Berzin D., Ivanyuk V. Practical application of clustering, classification and approximation methods based on neural networks. – 1st ed. – Moscow: Prometheus, 2020. – 448 p.

3. Dmitriev N.D. Clusterization of Russian companies in the technology sector based on market and financial indicators / N.D. Dmitriev, A.V. Tyulkova, V.V. Sorokozherdyev // Bulletin of the Chelyabinsk State University. – 2024. – № 10(492). – Pp. 116-127.

4. Tyukhova E.M. The use of cluster analysis for the formation of a securities portfolio in investment systems (roboadvisors) / E.M. Tyukhova, D.S. Sizykh // Management of large-scale systems development (MLSD'2019): Proceedings of the twelfth International conference, Moscow, October 01-03, 2019 / Under the general editorship of S.N. Vasiliev, A.D. Tsvirkun. Moscow: International Research Institute of Management Problems of the Russian Academy of Sciences, – 2019. – pp. – 210-220.

5. Broad Market Index. URL: <https://www.moex.com/ru/index/MOEXBMI/about> (accessed: 05/24/2024).

6. Zhdanov I.Y. Investment assessment of projects and business: a textbook / I.Y. Zhdanov, V.Y. Zhdanov. Moscow: Prospekt Publ., – 2019, – 120 p.

7. Borisova L.V., Sagaeva I.D. Models of optimal investment: an educational and methodical manual, Saratov. URL: http://elibrary.sgu.ru/uch_lit/1347.pdf (accessed: 05/23/2025).

8. Clusterization and partial learning. URL: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf> (accessed: 05/22/2025).

9. Takhir M.A., Ahmed A.U., Egorova L.D. On algorithms for the K-medoid problem // Actual problems of aviation and cosmonautics. – 2021. – No. 1. – pp. 194-196.

10. Dekanova K.V., Gate A.A. Methodology for forming an investment portfolio taking into account the individual characteristics of a private investor // Economics: yesterday, today, tomorrow. – 2023. – Volume 13. – No. 7A. – pp. 311-329.

11. Nourahmadi M., Sadeqi H. Portfolio Diversification Based on Clustering Analysis // Iranian Journal of Accounting, Auditing and Finance. – 2023. – №7(3) – P. 1-16.

12. Szczepocki, Piotr Clustering companies listed on the Warsaw Stock Exchange according to time-varying beta // Econometrics. – 2023. – №23 – P. 63-79.

13. Lim T., Ng H. Financial Portfolio Management Based on Shaped-Based Unsupervised Machine Learning: A Dynamic Time Warping Baycenter Averaging Approach to International Markets and Periods of Downside Event Risks // The Journal of Investing. – 2023. – №32. – P. 74-96.

14. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. – 12th ed. – Springer, 2017. – 745 p.

15. Fama, E. F., & French, K. R. The Cross-Section of Expected Stock Returns // Journal of Finance. – 1992. – №47(2). – P. 427-465.

16. Asness C.S., Moskowitz T., Pedersen L. Value and Momentum Everywhere // Journal of Finance. – 2013. – 68. – P. 929-985.

Назаров Тимур Рафикович, аспирант кафедры информатики, Российский экономический университет им. Г.В. Плеханова, г. Москва, Россия

Nazarov Timur Rafikovich, Postgraduate Student of the Department of Computer Science, Plekhanov Russian University of Economics, Moscow, Russia