

Экономико-математические методы и модели Economic & mathematical methods and models

Научная статья

УДК 330.4

DOI: <https://doi.org/10.18721/JE.18109>

EDN: <https://elibrary/VKOTSG>



АДАПТАЦИЯ K-MEANS КАК СРЕДСТВА АВТОМАТИЗАЦИИ ПРОЦЕССА ПРОГНОЗИРОВАНИЯ СЛАБОСТРУКТУРИРУЕМЫХ ВРЕМЕННЫХ РЯДОВ ЭКОНОМИЧЕСКОЙ ДИНАМИКИ

Л.К. Дунская , Е.В. Попова 

Кубанский государственный аграрный университет имени И.Т. Трубилина,
г. Краснодар, Российская Федерация

 dunskai.l@edu.kubsau.ru

Аннотация. В условиях растущего объема данных и увеличения сложности экономических взаимодействий возникает необходимость в применении более совершенных методов анализа и междисциплинарных подходов к исследованию систем со смешанным поведением. Интеллектуальные методы анализа, применяемые при машинном или глубоком обучении, позволяют учитывать сложные паттерны и нелинейные зависимости в данных. Методы прикладной статистики предоставляют надежные подходы к проверке гипотез, оценки параметров моделей и интерпретации результатов. При рассмотрении разных систем со сложным поведением выявлено, что экономические процессы часто характеризуются нелинейностью, нестационарностью и наличием скрытых зависимостей. Кроме того, применение методов машинного обучения и глубокого анализа данных позволяет не только повысить точность прогнозов, но и выявить скрытые закономерности, которые могут быть упущены при использовании традиционных статистических подходов. Это особенно важно при исследовании финансовых рынков, где динамика изменений может быть крайне нестабильной и подверженной влиянию множества внешних факторов. Внедрение таких методов способствует повышению эффективности принятия решений в условиях неопределенности, что делает их незаменимыми инструментами для современных экономических исследований. Таким образом, исследования в данной области являются актуальным направлением, что подтверждается не только природой ряда, но и необходимостью поиска более совершенных методов анализа и прогнозирования. В статье приводится предварительный анализ, а также построение прогноза на базе линейного клеточного автомата. Методы прикладной статистики и интеллектуального анализа данных выступают инструментами в части анализа временного ряда, а также применяются в вопросе адаптации методов кластеризации как средства автоматизации прогнозной модели. Использование и встраивание в алгоритм линейного клеточного автомата известных методов кластеризации позволяют, как показано авторами, выявить закономерности и повысить качество прогноза. Объектом исследования является временной ряд финансового рынка, поскольку данные экономические ряды демонстрируют влияние множества сложновывявляемых (по степени воздействия) факторов, таких как внешние шоки, сезонные колебания и долгосрочные тренды. По результатам исследования выяснено, что использование алгоритмов интеллектуального анализа позволяет автоматизировать процесс перевода числовых показателей временного ряда в лингвистический аналог для получения прогнозного значения без потери качества.

Ключевые слова: прогноз, интеллектуальный анализ данных, методы прикладной статистики, линейный клеточный автомат

Для цитирования: Дунская Л.К., Попова Е.В. (2025) Адаптация k-means как средства автоматизации процесса прогнозирования слабоструктурируемых временных рядов экономической динамики. П-Economy, 18 (1), 160–177. DOI: <https://doi.org/10.18721/JE.18109>

Research article

DOI: <https://doi.org/10.18721/JE.18109>

ADAPTATION OF K-MEANS TO AUTOMATED FORECASTING OF POORLY STRUCTURED TIME SERIES OF ECONOMIC DYNAMICS

L.K. Dunskaja  , E.V. Popova 

Kuban State Agrarian University named after I.T. Trubilin,
Krasnodar, Russian Federation

✉ dunskaja.l@edu.kubsau.ru

Abstract. With the growing volume of data and increasing complexity of economic interactions, more advanced analysis methods and interdisciplinary approaches should be applied to study of systems with mixed behavior. Data mining methods used in machine learning or deep learning allow to take into account complex patterns and nonlinear dependencies in the data. Applied statistics methods provide reliable approaches to hypothesis testing, model parameter estimation and interpretation of results. It was established for different systems with complex behavior that economic processes are often characterized by nonlinearity, instability, and the presence of hidden dependencies. Furthermore, machine learning and deep data analysis methods allow not only to improve the forecasting accuracy but also to identify hidden patterns that may be overlooked by traditional statistical approaches. This is especially important in the study of financial markets, where the dynamics of change can be extremely unstable and influenced by many external factors. Such methods help to increase the effectiveness of decision-making in conditions of uncertainty, serving as indispensable tools for modern economic research. Thus, research in this area is urgent, as confirmed not only by the nature of the series, but also by the need to find more advanced methods of analysis and forecasting. The article provides preliminary analysis, additionally constructing a forecast based on a linear cellular automaton. Applied statistics and data mining tools were used for time series analysis as well as for adaptation of clustering methods as a means for automating the predictive model. We confirmed that use and integration of well-known clustering methods into the linear cellular automaton algorithm allows to identify patterns and improve the quality of the forecast. The object of the study is the time series of the financial market, since these economic series are influenced by a variety of factors that are hard to detect (in terms of their influence), such as external shocks, seasonal fluctuations and long-term trends. Our findings indicate that data mining algorithms make it possible to automate the process of translating numerical indicators of a time series into a linguistic equivalent to obtain predictive values without loss of quality.

Keywords: forecasting, data mining, methods of applied statistics, linear cellular automaton

Citation: Dunskaja L.K., Popova E.V. (2025) Adaptation of k-means to automated forecasting of poorly structured time series of economic dynamics. *П-Economy*, 18 (1), 160–177. DOI: <https://doi.org/10.18721/JE.18109>

Введение

Актуальность

При появлении новых вычислительных мощностей, технологии интернета вещей и множества устройств, способных собирать данные в реальном времени, мир исследователей столкнулся с таким феноменом как большие данные (Big Data). Объемы информации растут в геометрической прогрессии, а массив информации, собираемый в рамках одной системы, становится все более разнообразным за счет фиксации наблюдений по новым показателям. Это говорит о том, что анализ систем становится более сложным и требует применения новых методов моделирования и прогнозирования.

Современные системы сами по себе демонстрируют сложное поведение, которое необходимо учитывать в исследованиях для получения более достоверных прогнозов. Так, например, для стохастических систем характерны элементы случайности, выражающиеся в непредсказуемом

поведении, для динамических систем – периодические или сезонные изменения по определенным законам, а для детерминированных систем возможно прогнозирование будущих состояний при определенно заданных начальных условиях¹.

В рамках данного исследования рассмотрен подход к анализу и прогнозированию системы, демонстрирующей сложное и многогранное поведение, связанное со специфической природой временного ряда и зависимостью от множества внутренних и внешних факторов воздействия [1, 2]. Работа включает предпрогнозный анализ ряда, а также комбинацию инструмента прогнозирования с методом кластеризации данных для выявления скрытых паттернов поведения.

Литературный обзор

В исследовании проводится прогнозирование по временному ряду – комплексу зафиксированных через равные промежутки времени состояний системы². Авторами предлагается провести работу над измерениями системы, показывающую стохастическую природу, реакцию на мало выявленные факторы, зависящие от исторических и социальных событий – экономический временной ряд финансового рынка [3].

Прикладные статистические исследования играют ключевую роль в анализе, прогнозировании и моделировании сложных систем. В контексте анализа складывающейся конъюнктуры и перспективных вариантов развития ученые выявляют закономерности, измеряют ключевые параметры и в дальнейшем строят прогнозы на основе изменения данных параметров с целью разработки сценариев оптимального управления системой в целом. Таким образом, минимальное исследование экономических рядов включает определение сезонной и трендовой компонент³. Данные компоненты являются основными, поскольку сезонность выявляет предсказуемые колебания, а тренд показывает долгосрочное направление изменения данных, что необходимо учитывать при построении прогноза [4].

Инструментом из области математического моделирования, связанным с теорией сложных систем, является линейный клеточный автомат. Несмотря на сложность прогнозирования описываемых систем, данный инструмент показывает возможность построения прогноза для систем любого вида при малой ошибке прогноза [5]. Подробное описание алгоритма прогнозирования линейным клеточным автоматом представлено в работе [6].

Несмотря на то, что линейный клеточный автомат применим в построении прогнозов разных систем, при его построении есть шаги, которые сложны в автоматизации программными средствами. Такие шаги требуют ручного ввода параметров от эксперта, который строит прогноз. В рамках данной работы таким шагом является перевод числовых показателей в лингвистический временной ряд.

Перевод происходит следующим образом:

1) определение множества термов, которые характеризуют каждое значение временного ряда. Оптимальным множеством считается $\bar{U} = \{H, C, B\}$, где H – низкий, C – средний и B – высокий, определяющие низкие, средние и высокие значения соответственно;

2) перевод значения в лингвистические термы согласно множеству \bar{U} . Вручную данный этап происходит через шаги, которые задает исследователь, например:

а) выбор максимальных и минимальных значений ряда, определенных на всем протяжении графика, т.е. в процессе визуального анализа эксперт определяет такие точки на основе поведения временного ряда во времени;

б) построение линий по выбранным точкам, которые называются НОЛ и ВОЛ – нижняя и верхняя огибающие линии (соответственно). Данные линии на графике выглядят как коридор и, в свою очередь, делятся на три дорожки, в которые попадают значения ряда и определяются

¹ Чернышов В.Н., Чернышов А.В. (2008) *Теория систем и системный анализ: учебное пособие*, Тамбов: Изд-во Тамб. гос. техн. ун-та.

² Афанасьев В.Н., Юзбашев М.М. (2001) *Анализ временных рядов и прогнозирование: учебник*, М.: Финансы и статистика.

³ Музыка О.А. (2011) Бифуркации в природе и обществе: естественнонаучный и социосинергетический аспект. *Современные наукоемкие технологии*, 1, 87–91.



в соответствии с термами (средние, низкие и высокие значения). Таким образом, значения, расположенные ниже коридора или попадающие в зону нижних значений, определяются термом H ; значения, находящиеся в центральной зоне коридора, соответствуют средним значениям и маркируются термом C ; остальным значениям, входящим в верхнюю зону коридора или находящимся выше него, присваивается значение B .

Описанный выбор точек для построения коридора значений не является исключительно верным для каждого ряда. Исследования показали, что необходимо применять разные подходы для построения ВОЛ и НОЛ [7].

Авторами предлагается рассмотреть методы интеллектуального анализа данных в части распределения значений по подмножествам термов [8–11]. Обоснованием выбора инструментов данной области является следующее: если в массиве данных присутствует информация о скрытых влияниях на систему, то они могут быть выявлены в рамках кластеризации, поскольку данный подход используется в рамках анализа больших данных именно с этой целью [12–14].

Исходя из актуальности и литературного обзора, целью исследования является разработка комбинированного подхода к прогнозированию временных рядов, основанного на интеграции линейного клеточного автомата с методом кластеризации k -means (метод k -средних) для повышения точности и интерпретируемости предсказаний в сложных системах [15–19].

В рамках работы поставлены следующие задачи:

- 1) исследование динамических свойств системы путем определения сезонности и тренда методами автокорреляции и скользящим средним;
- 2) подготовка ряда к процессу кластеризации методом k -means;
- 3) проведение кластеризации через определение евклидова расстояния между векторами признаков;
- 4) перевод показателей в лингвистический ряд на основе результатов кластеризации;
- 5) построение прогноза с использованием линейного клеточного автомата с применением кластеризации и через ручной ввод точек границ коридора значений.

Объектом исследования является финансовый временной ряд СПБ Биржи по ценам закрытия месяца акций компании Cisco.

Предметом исследования выступают статистические методы, модель построения прогноза систем смешанного поведения и метод интеллектуального анализа данных – кластеризация.

Методы и материалы

Первый этап исследования посвящен анализу выявления такого свойства, как тренд-сезонность временного ряда, с использованием методологии расчета скользящих средних и автокорреляции.

Автокорреляция (ACF) – мера зависимости между значениями временного ряда на разных временных интервалах, называемых шагами или лагами (k) [17, 18]. Уравнение автокорреляции представлено в формуле (1).

Исходя из определения, автокорреляция позволяет вычислить то, на сколько текущие значения временного ряда зависят от предыдущих. Таким образом, можно выявить скрытые зависимости, которые влияют на наличие тренда и сезонности в определенные периоды исследуемой системы. Интерпретация результатов происходит при сравнении полученного динамического показателя по его нахождению относительно 0. Высокая автокорреляция на больших лагах показывает наличие растущего тренда. С точки зрения сезонности чем выше автокорреляция, тем большую зависимость имеет ряд в конкретный исследуемый шаг.

$$ACF(k) = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}, \quad (1)$$

где x_t – значение временного ряда в момент времени t , \bar{x} – среднее значение временного ряда, k – лаг (разница во времени между наблюдениями), T – общее количество наблюдений в ряду.

Скольльзящая средняя – метод сглаживания временного ряда, при котором выявляются долгосрочные тенденции и сезонные колебания [20–23]. В рамках исследования рассмотрены два типа скользящих, представленных в уравнениях (2) и (3) соответственно:

– простая скользящая средняя (*SMA*) – вычисляется среднее значение временного ряда за определенный период (t);

– взвешенная скользящая средняя (*WMA*) – вычисляется как сумма взвешенных значений временного ряда, где каждому значению присваивается вес (ω) в зависимости от его значимости. Эти веса могут быть различны в рамках одного периода времени (t). Затем эта сумма делится на сумму весов для нормализации значений:

$$SMA_t = \frac{1}{n} \sum_{i=t-n+1}^t x_i, \quad (2)$$

$$WMA_t = \frac{\sum_{i=t-n+1}^t \omega_i x_i}{\sum_{i=t-n+1}^t \omega_i}, \quad (3)$$

где x_i – значение временного ряда в момент времени i , n – размер так называемого окна – определенное и фиксированное количество точек, ω – вес, присвоенный соответствующему значению.

Вторым этапом исследования является подготовка временного ряда к процессу кластеризации. Поскольку методы кластеризации при работе с временными рядами сами по себе не могут выявлять зависимость таких периодов, как кварталы, года и т.д., и соотносить их с поведением ряда, необходимо добавить критерии по их идентификации [23].

Также необходимо учитывать, что экономические временные ряды часто характеризуются высокой волатильностью и наличием случайных процессов, а ряды по состоянию финансовых рынков часто показывают нелинейные и нестационарные характеристики. Таким образом, в качестве дополнительного определения характера значений ряда с целью распределения их на кластеры по сезонным и трендовым компонентам необходимо внедрить несколько дополнительных показателей.

Учитывая сложность интерпретации финансовых рядов, авторы предлагают добавить дополнительную шкалу оценки каждого значения по поведенческой принадлежности показателей ряда в зависимости от сезона и сезонного тренда. Принятая оценка по шкале поведение-сезон выражена числовым значением для каждой строки в закрытом полуинтервале $[0, K]$, где K – количество показателей шкалы.

При комплексном использовании результатов первого этапа и полученных выводов о показателе автокорреляции идентификатор части «сезон» основывается на результатах первого шага исследования, а критерий «поведение» выявляется благодаря визуальному анализу по графику временного ряда и определению отрезков роста и спада значений.

В рамках гармонического анализа разложение ряда на синусоидальные и косинусоидальные компоненты призвано учесть регулярные колебания, например годовые или квартальные сезонные эффекты, экономические циклы и т.п. [21, 22]:

$$x_{\sin e} = \sin\left(\frac{2\pi x}{N}\right), \quad (4)$$

$$x_{\cos e} = \cos\left(\frac{2\pi x}{N}\right), \quad (5)$$



где $x_{\sin e}$ – синусоидальное представление значения ряда, $x_{\cos e}$ – косинусоидальное представление значения ряда, x – значение ряда, N – количество значений во временном ряду.

По итогу второго этапа исследования временной ряд в виде [«Дата», «Показатель»] преобразуется в массив данных с содержанием [«Дата», «Показатель», «Поведение-сезон», «Синус-компонента», «Косинус-компонента»].

В дальнейших расчетах используется матрица размерности 62×4 , где каждая строка представляет собой вектор, описывающий показатель и три его характеристики:

$$A = \{a_i = (x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}) | i = 1, \dots, 62\}, \quad (6)$$

где a_i – i -й вектор строки, $x_{i,1}$ – показатель, $x_{i,2}$, $x_{i,3}$, $x_{i,4}$ – характеристики показателя (поведение-сезон, синус- и косинус-компоненты).

Третий этап исследования посвящен кластеризации данных. Интеллектуальный анализ данных предоставляет различные модели для проведения этого этапа. Модели кластеризации имеют принципиальные различия не только в том, для данных какой области они применяются, а также в том, какой алгоритм или подход заложен в процесс разбиения данных по кластерам. В рамках исследования принято решение рассматривать методы, основанные на плотности распределения, связанности данных и методы, алгоритмы которых связаны с определением расстояний.

В качестве метода на основе плотности распределения, применяемого к финансовым рядам, рассматривался DBSCAN. В результате кластеризации эксперту предоставляется результат в виде двух типов объектов: кластеры, характеризующиеся высокой плотностью значений по заданным параметрам; шум – области, которые демонстрируют низкую плотность распределения, когда конкретные точки не принадлежат ни одному кластеру. Для настройки модели необходимо предварительно вывести такие обязательные параметры, как радиус окрестности поиска соседних точек и минимальное количество точек в окрестности выявленного радиуса. Можно говорить о том, что основной целью кластеризации можно считать выявление кластеров произвольной формы и эффективной фильтрации шума [24, 25]. Данное направление не связано с задачами текущего исследования, но может использоваться в работе по уменьшению ошибки прогнозного значения в рамках дальнейших исследований.

С точки зрения исследования связанности данных рассматривался метод Agglomerative Clustering. В результате работы алгоритма эксперту предоставляется дендрограмма – вид диаграммы, демонстрирующий древовидное распределение данных по кластерам, где по горизонтальной оси располагаются точки данных, а по вертикальной приводится мера расстояний между кластерами. Визуальный анализ дендрограммы дает возможность показать, какое число кластеров позволит оптимально разбить данные и сформировать группы по наибольшему уровню схожести, обобщенно – выявить иерархическую зависимость между объектами [26, 27]. Данное исследование также не преследует цели этой работы, смещая фокус внимания на поиск оптимального числа кластеров, характеризующих схожесть объекта, и глубокое исследование взаимосвязей данных в принципе. Исследование в части оптимального числа кластеров для уменьшения ошибки прогноза тоже может быть проведено в рамках последующих работ.

Рассматривая приведенные методы кластеризации, авторы выделили определенные критерии для поиска:

– Доступность интерпретации математических уравнений кластерного анализа для поддержания прозрачности вычислительного процесса. Этот критерий говорит о том, что упор делается не в сторону поиска подходящих кластеров, а в сторону прозрачности научного исследования в части возможности практического применения кластеризации в части малых временных срезов данных;

– Исходя из доступности интерпретации, необходимым моментом будет считаться факт уменьшения вероятности появления ошибок в результате расчета без использования программных продуктов. Например, при проведении иерархической кластеризации необходимо проводить процесс последовательного вычисления метрик расстояний между всеми парами объектов в пространстве признаков, а также интеграции этих расстояний для формирования новой иерархической структуры классов на каждой возможной итерации. Вручную такие итерации занимают много времени, а кроме того, приводят к появлению других субъективных факторов, что влечет потерю качества исследования.

– Модель должна поддерживать анализ объектов по временным рядам небольшой протяженности. Исходя из концепции того, что временные ряды отражают информацию как о действии системы, так и о ее реакции на влияние окружающего мира, исключение шумов и аномалий не рассматривается.

Таким образом, в рамках данного исследования выбран метод k-means. Данный метод характеризуется простотой и прозрачностью расчетов, способностью быстро и эффективно работать с компактными кластерами. Модель не требует тонкой настройки параметров, принимая на вход только число и характеристики центроидов. Исследуемый ряд показывает наличие сезонности и выраженного тренда, что метод кластеризации может адекватно обработать.

Принимая во внимание специфику построения прогноза линейным клеточным автоматом с использованием распределения значений по трем лингвистическим термам, соответствующим высоким, средним и низким значениям, авторы приняли решение кластеризовать данные вокруг трех центроидов (распределить по трем кластерам).

Процесс кластеризации состоит из итерационного повторения следующих шагов [28–30]:

- 1) инициализация (определение) центроидов кластеров – средних значений всех точек в соответствующем кластере (C_1, C_2, C_3);
- 2) распределение точек по кластерам:
 - а) расчет евклидова расстояния ($d_{C_m,j}$) между показателем ($x_{j,1}$) и его характеристиками ($x_{i,2}, x_{i,3}, x_{i,4}$) по j -й строке и соответствующим средним значением каждого кластера ($c_{m,j}$):

$$d_{C_m,j} = \sqrt{(x_{i,1} - c_{m,1})^2 + (x_{i,2} - c_{m,2})^2 + (x_{i,3} - c_{m,3})^2 + (x_{i,4} - c_{m,4})^2}; \quad (7)$$

- б) Присвоение точки тому кластеру, центр которого находится ближе всего к ней. Для этого при сравнении расстояний бралось минимальное значение из трех;

- 3) пересчет центроидов кластеров – рассчитывается среднее значение всех точек, входящих в кластер.

Выходом из цикла служит проверка дельт показателей центроидов в виде $|C_i^{t+1} - C_i^t|$ на минимальное отклонение. В данной формуле t – обозначение центра кластера предыдущей итерации, $t + 1$ – обозначение центра кластера в рассчитываемой (новой) итерации. За минимально значимое изменение (tol) принят порог, обозначающий изменение, равный не менее 0,01%. Такими порогами стали значения пересчитанных центров кластера после первой итерации, определенные на уровне 1×10^{-4} . С учетом того, что в расчетах присутствуют компоненты синуса и косинуса, авторами произведено сравнение с использованием знака нечеткого равенства, т.е. по формуле $|C_i^{t+1} - C_i^t| \leq tol$.

В части перехода к построению прогноза линейным клеточным автоматом принято решение построить две модели: на основе ручного перевода лингвистических переменных и с автоматизацией данного шага через применение кластеризации. Первый вариант, удовлетворяющий критерию достоверности (малой ошибке прогноза), использован в качестве эталона при сравнении результатов.



Оценка построенного прогноза определяется путем вывода отношения неугаданных значений к общему количеству точек ряда в процентном соотношении [31, 32].

Адаптация методов кластеризации считается успешной при условии получения ошибки прогноза менее показателя ошибки эталонной модели или не более 25%.

Объектом исследования выступает финансовый ряд ежемесячных значений закрытия цены⁴ по акциям Cisco СПб Биржи за период с апреля 2013 г. по май 2018 г.

Результаты и обсуждения

На первом этапе исследования необходимо определить параметры лага для расчета автокорреляции. Поскольку временной ряд состоит из фиксации значений по месяцам, предлагается рассматривать шаги, равные 2, 3, 6 и 12 месяцам. Графики простой скользящей средней с шагами 2, 3, 6 и 12 приведены на рис. 1.

При построении скользящего среднего по предложенным лагам выявлено, что шаг в 2 месяца показывает достаточно выраженное колебание, что говорит о недостаточном сглаживании значения, в то время как при шагах в 6 и 12, напротив, видно избыточное усреднение значений. Таким образом, в промежутке между лагами 3 и 6 происходит потеря влияния прошлых значений на будущее. В рамках данного предположения принято выявить дополнительный коэффициент автокорреляции для шага в 4 месяца. В табл. 1 приведены рассчитанные коэффициенты для исследуемого ряда.

Таблица 1. Коэффициенты автокорреляции для финансового ряда ежемесячных показателей закрытия цен на акции компании Cisco

Table 1. Autocorrelation coefficients for financial series of monthly stock closing prices for Cisco

№	Лаг (число месяцев)	Показатель
1	2	0,538
2	3	0,409
3	4	0,197
4	6	0,063
5	12	-0,297

По результатам определения коэффициентов видно, что лаги 2, 6 и 12 недостаточно выявляют динамические особенности ряда.

Рассчитанные дельты для лагов под номерами 1–2, 2–3 и 3–4 дают значения в 0,129, 0,212 и 0,134 соответственно. Самый большой показатель разности между значениями зафиксирован для лага, равного шагу в 3 месяца (0,212), что говорит о выраженной сезонности по кварталам. Этот вывод является результатом первого этапа работы.

На втором этапе исследования выведем шкалу, по которой будет возможен процесс присвоения каждому значению ряда характеристики в рамках поведение-сезон, а также модифицируем временной ряд в матрицу показателей и векторов его характеристик.

В данном исследовании в качестве поведения экономического ряда приняты обозначения: спад – для определения тенденции к падению цены, рост – для определения роста этого показателя. Для фиксации периодов конкретного поведения построен график по усредненным значениям скользящей средней для лага 6, вертикальными линиями обозначены точки изменения поведения. Полученное изображение для анализа приведено на рис. 2.

⁴ Источник данных: www.finam.ru.

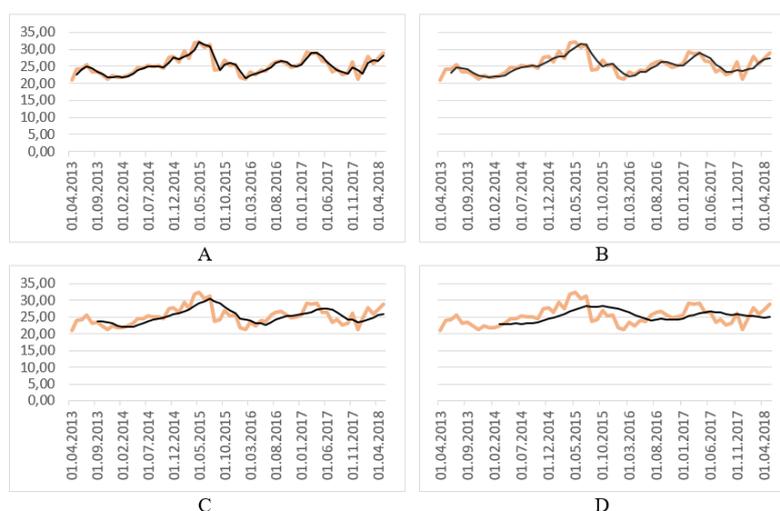


Рис. 1. Показатели автокорреляции по финансовому ряду цен акций Cisco с лагом:

A – 2 месяца, B – 3 месяца, C – 6 месяцев, D – 12 месяцев

Fig. 1. Autocorrelation indicators for financial series of Cisco stock prices with lags:

2 months (A), 3 months (B), 6 months (C), 12 months (D)

В результате визуального анализа представлено следующее поведение системы: до 01.03.2014 – спад, до 01.07.2015 – рост, до 01.04.2016 – спад, до 01.04.2017 – рост, до 01.01.2018 – спад, остальные значения – рост.

Учитывая результаты предыдущего шага о влиянии сезонности на ряд, в части определения «сезона» в показателе поведение-сезон принято использовать множество вида {весна, лето, осень, зима}.

После построения лингвистических определений в формате поведение-сезон каждой строке присвоен числовой показатель в диапазоне положительных значений целых чисел. Результат построения шкалы характеристики показателей приведен в табл. 2.

Таблица 2. Характеристики значений ряда с учетом поведения и сезонной составляющей в виде лингвистического и числового показателей
Table 2. Characteristics of series values accounting for behavioral and seasonal components as linguistic and numerical indicators

Определение вида поведение-сезон	Числовое представление
рост весна	0
спад весна	1
рост осень	2
спад осень	3
рост зима	4
спад зима	5
рост лето	6
спад лето	7

В качестве дополнительной идентификации динамических характеристик произведено преобразование дискретного значения по столбцу показателей закрытия цены месяца путем добавления синусоидальных и косинусоидальных компонент, рассчитанных по формулам (4) и (5).

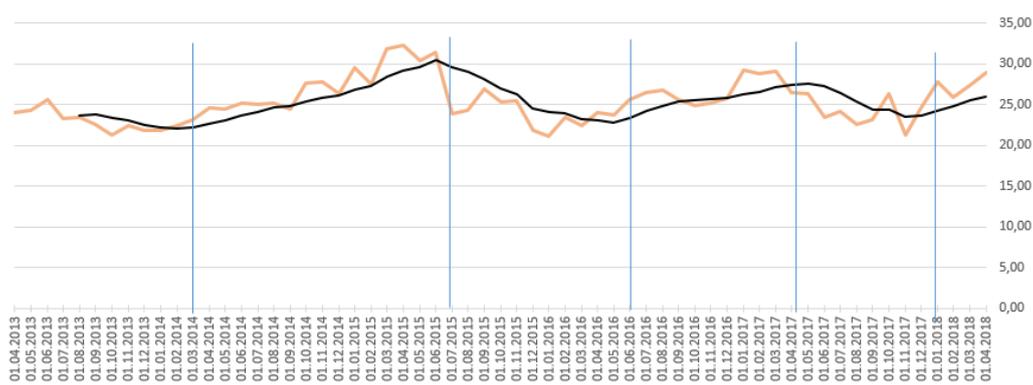


Рис. 2. Определение периодов роста и спада временного ряда цен закрытия на акции по скользящей средней с лагом 6
 Fig. 2. Identification of growth and decline periods in time series for stock closing price based on 6-month lag moving average

Итоговая таблица значений с результатами преобразования ряда динамического показателя в числовое представление поведение-сезон и расчета компонент синуса и косинуса представлены в столбцах 2–5 в табл. 4.

В рамках кластеризации расчет по столбцам дат проводиться не будет.

Переходим к построению прогноза по «классическому» переводу временного ряда в соответствующие термы через определение коридора значений. Для наглядности столбцы раскрашены в соответствии с принадлежностью к множеству низких (красные), средних (желтые) и высоких (зеленые) значений. Результат раскраски ряда представлен на рис. 3.

При переходе к этапу кластеризации предполагается использование усредненных значений центроидов, которые должны отображать логическое соответствие термам $\bar{U} = \{H, C, B\}$.

В рамках определения начальных центров взяты показатели по возрастанию, исходя из множества всех значений по каждому столбцу, таким образом, чтобы кластеры C1, C2 и C3 описывали низкие, средние и высокие значения соответственно.

По итогам проведения циклических расчетов методом k-means выявлено отсутствие минимально значимых изменений на 6-й итерации.

Начальные и нормализованные (после обновления) центры кластеров приведены в табл. 3.

Таблица 3. Начальные и нормализованные значения центроидов кластеров C1, C2, C3
Table 3. Initial and normalized values of centroids for clusters C1, C2, C3

	Центроид	Закрытие цены	Число	Синус	Косинус
Начальные значения центроидов	C1	20	2	0	0
	C2	23	3	0,7	0,7
	C3	29	6	1	1
Окончательная 6-итерация пересчета центроидов	C1	23,68	1,70	0,62	0,13
	C2	24,14	5,71	-0,69	-0,15
	C3	28,67	2,17	0,12	0,13

Итоговое распределение точек по соответствующим кластерам приведено в табл. 4, столбец «Принадлежность к кластеру». Для дальнейшего построения прогноза произведен перевод показателей временного ряда в лингвистический аналог в зависимости от принадлежности к кластеру, где C1 соответствует терму H, C2 – терму C, C3 – терму B. Результат присвоения соответствующего терма приведен в табл. 4, столбец «Терм».

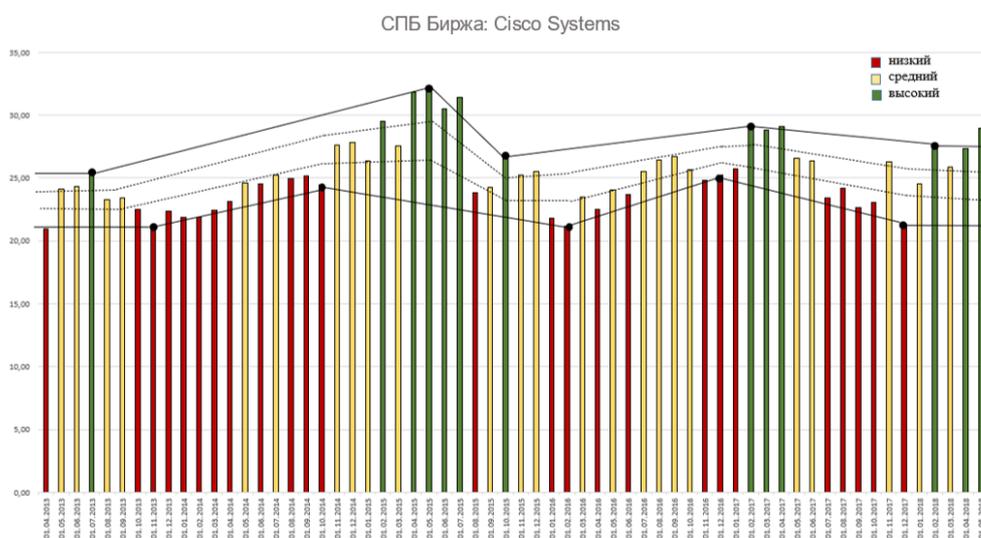


Рис. 3. Гистограмма перевода ряда в лингвистический аналог через использование коридора значений.

Принадлежность точек к соответствующему множеству определено цветом

Fig. 3. Histogram for translation of the series to linguistic equivalent using a value corridor.

Colors of the points correspond to the sets that the points belong to

На рис. 4 представлена гистограмма, на которой разными принтами обозначена принадлежность показателя к определенному кластеру, где темным цветом закрашены низкие значения, белым – высокие, значения с принтом – средние.

Построение прогноза линейным клеточным автоматом через «классическое определение» лингвистического временного ряда дало следующие результаты: ошибка прогноза равна 19%; прогноз говорит о том, что показатель ожидается высоким (соответствует терму В) и будет приближен к значению 26,52.

Результат построения прогноза через внедрение методов кластеризации: ошибка 15%; построенный прогноз говорит о принадлежности к средним значениям с приближением к показателю в 25,23.

Заключение

Отдельно отметим, что получение результата со значительным улучшением уровня ошибки в вероятностной оценке на предложенной авторами малой выборке подтверждает гипотезу о необходимости дальнейшего исследования возможностей кластеризации при обучении линейного клеточного автомата в направлении увеличения числа кластеров, не ограничиваясь тремя кластерами.

Таким образом, анализ полученных результатов выявил, что предложенные к встраиванию в модель линейного клеточного автомата методы кластеризации не только позволяют автоматизировать этап, который ранее требовал участия эксперта или лица, принимающего решение, но и существенно улучшить значение ошибки прогноза. Таким образом подтверждается:

- целесообразность и возможность встраивания методов кластеризации в методы прогнозирования нелинейных систем;
- возможность автоматизации этапа «обучения» линейного клеточного автомата без участия эксперта;
- универсальность применения инструмента линейного клеточного автомата к прогнозированию сложных систем.

Таблица 4. Итог преобразования временного ряда в массив данных, содержащих показатели и векторы параметров, отражающие сезонные и поведенческие компоненты
Table 4. Time series transformed into data array containing indicators and parameter vectors reflecting seasonal and behavioral components

Дата	Показатель (цена закрытия)	Числовой параметр поведение-сезон	Синус-компонента	Косинус-компонента	Принадлежность к кластеру	Терм
1	2	3	4	5	6	7
01.04.2013	20,93	1	0,852	-0,523	C1	H
01.05.2013	24,09	1	0,644	-0,765	C1	H
01.06.2013	24,30	7	0,628	-0,778	C2	C
01.07.2013	25,56	7	0,524	-0,852	C2	C
01.08.2013	23,31	7	0,703	-0,711	C2	C
01.09.2013	23,43	3	0,694	-0,720	C1	H
01.10.2013	22,51	3	0,758	-0,652	C1	H
01.11.2013	21,25	3	0,835	-0,550	C1	H
01.12.2013	22,40	5	0,765	-0,644	C2	C
01.01.2014	21,89	5	0,798	-0,603	C2	C
01.02.2014	21,84	5	0,801	-0,599	C2	C
01.03.2014	22,41	1	0,765	-0,644	C1	H
01.04.2014	23,11	0	0,717	-0,697	C1	H
01.05.2014	24,62	0	0,602	-0,798	C1	H
01.06.2014	24,52	6	0,611	-0,792	C2	C
01.07.2014	25,24	6	0,551	-0,834	C2	C
01.08.2014	24,98	6	0,573	-0,820	C2	C
01.09.2014	25,18	2	0,556	-0,831	C1	H
01.10.2014	24,45	2	0,616	-0,788	C1	H
01.11.2014	27,65	2	0,333	-0,943	C3	B
01.12.2014	27,81	4	0,318	-0,948	C3	B
01.01.2015	26,37	4	0,452	-0,892	C2	C
01.02.2015	29,51	4	0,150	-0,989	C3	B
01.03.2015	27,52	0	0,345	-0,938	C3	B
01.04.2015	31,83	0	-0,084	-0,996	C3	B
01.05.2015	32,32	0	-0,133	-0,991	C3	B
01.06.2015	30,47	6	0,054	-0,999	C3	B
01.07.2015	31,43	6	-0,044	-0,999	C3	B
01.08.2015	23,87	7	0,661	-0,750	C2	C
01.09.2015	24,24	3	0,633	-0,774	C1	H
01.10.2015	26,87	3	0,406	-0,914	C3	B
01.11.2015	25,27	3	0,549	-0,836	C1	H
01.12.2015	25,53	5	0,526	-0,850	C2	C
01.01.2016	21,78	5	0,804	-0,594	C2	C
01.02.2016	21,18	5	0,839	-0,544	C2	C
01.03.2016	23,48	1	0,690	-0,723	C1	H

Окончание таблицы 4

Дата	Показатель (цена закрытия)	Числовой параметр поведение-сезон	Синус- компонента	Косинус- компонента	Принадлежность к кластеру	Терм
1	2	3	4	5	6	7
01.04.2016	22,48	1	0,760	-0,650	C1	H
01.05.2016	24,04	0	0,648	-0,761	C1	H
01.06.2016	23,69	6	0,675	-0,738	C2	C
01.07.2016	25,55	6	0,525	-0,851	C2	C
01.08.2016	26,44	6	0,446	-0,895	C2	C
01.09.2016	26,72	2	0,420	-0,907	C3	B
01.10.2016	25,68	2	0,513	-0,858	C1	H
01.11.2016	24,85	2	0,584	-0,812	C1	H
01.12.2016	25,22	4	0,553	-0,833	C2	C
01.01.2017	25,72	4	0,510	-0,860	C2	C
01.02.2017	29,18	4	0,183	-0,983	C3	B
01.03.2017	28,79	0	0,222	-0,975	C3	B
01.04.2017	29,07	0	0,194	-0,981	C3	B
01.05.2017	26,56	1	0,435	-0,900	C3	B
01.06.2017	26,33	7	0,456	-0,890	C2	C
01.07.2017	23,45	7	0,693	-0,721	C2	C
01.08.2017	24,21	7	0,635	-0,772	C2	C
01.09.2017	22,63	3	0,750	-0,661	C1	H
01.10.2017	23,10	3	0,718	-0,696	C1	H
01.11.2017	26,31	3	0,458	-0,889	C3	B
01.12.2017	21,29	5	0,833	-0,554	C2	C
01.01.2018	24,57	5	0,606	-0,795	C2	C
01.02.2018	27,79	4	0,320	-0,948	C3	B
01.03.2018	25,85	0	0,499	-0,867	C1	H
01.04.2018	27,35	0	0,362	-0,932	C3	B
01.05.2018	28,93	0	0,208	-0,978	C3	B

Направление дальнейших исследований

Учитывая факт того, что при внедрении методов кластеризации в процесс «обучения» линейного клеточного автомата значение рассчитанной ошибки прогноза имеет тенденцию к уменьшению, нельзя не отметить следующее наблюдение: значительное смещение неугаданных значений в конец временного ряда.

Приведем пример сравнения лингвистических рядов, полученных: по результатам кластеризации; по результатам построения линейного клеточного автомата (рис. 5).

Такое поведение подтверждает вывод, что прогнозная модель имеет тенденцию к переобучению на ранних исторических данных временного ряда, что, в свою очередь, приводит к появлению большого числа неугаданных значений при прогнозировании более поздних по временной шкале показателей и влияет на ошибку достоверности прогноза.

В качестве дальнейшего исследования предлагается поставить цель в увеличении точности предсказания ряда (в случае наиболее близких показателей к прогнозным значениям) через следующие задачи и направления:

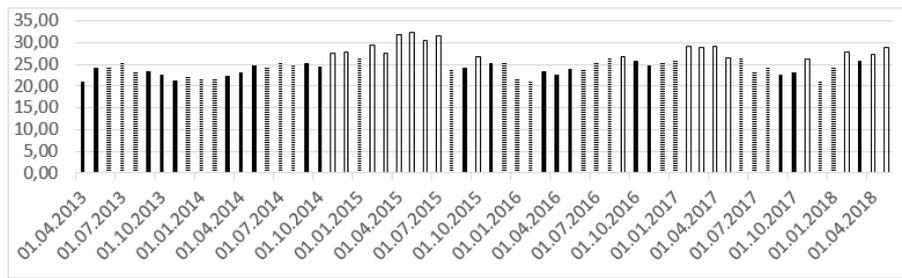


Рис. 4. Гистограмма с раскраской значений по результатам кластеризации
Fig. 4. Histogram with color-coded values based on clustering results

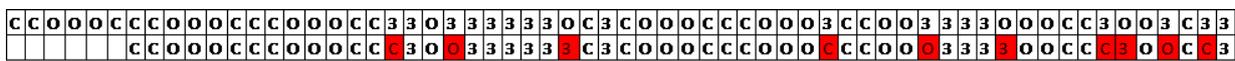


Рис. 5. Сравнение лингвистических рядов перевода ряда цен закрытия акций через кластеризацию (верхний)
и по результатам линейного клеточного автомата (нижний)

Fig. 5. Comparison of linguistic series for translation of stock closing price series obtained by clustering (top)
and linear cellular automaton (bottom)

- рассмотрение возможности использования методологии присвоения весов значимости значениям, расположенным ближе к концу временного ряда при анализе, кластеризации или построении прогноза;
- определение оптимального числа кластеров с наименьшей ошибкой прогноза;
- исследование аномалий и компонент временного ряда, фиксация которых переобучает модель.

СПИСОК ИСТОЧНИКОВ

1. Nicolis G. (1995) *Introduction to Nonlinear Science*, Cambridge, UK: Cambridge University Press.
2. Jagadev A.K., Routra N.K., Sarangi P.P., Patra S.K., Paikaray B.K. (in press) (2024) Fuzzy inventory modelling: addressing uncertainty in economic order quantity analysis within nonlinear science. *International Journal of Applied Nonlinear Science*. DOI: <https://doi.org/10.1504/ijans.2024.10066705>
3. Кумратова А.М., Попова Е.В., Савинская Д.Н., Курносоева Н.С. (2015) Комплексная методика анализа экономических временных рядов методами нелинейной динамики. *Современная экономика: проблемы и решения*, 8 (68), 35–43. DOI: <https://doi.org/10.17308/meps.2015.8/1292>
4. Савинская Д.Н., Великанова Л.О., Дунская Л.К., Попова М.И. (2020) Предпрогнозный анализ временного ряда с выявлением тренд-сезонных компонент методом Четверикова с использованием средств MS Excel. *Современная экономика: проблемы и решения*, 10 (130), 18–25. DOI: <https://doi.org/10.17308/meps.2020.10/2445>
5. Кумратова А.М., Попова Е.В., Костенко И.В., Дунская Л.К. (2019) Адаптация линейного клеточного автомата для решения задач прогнозирования на базе природно-экономических временных рядов. *Современная экономика: проблемы и решения*, 9 (117), 8–17. DOI: <https://doi.org/10.17308/meps.2019.9/2196>
6. Кумратова А.М., Попова Е.В., Биджиев А.З. (2014) *Экономико-математическое моделирование риска в задачах управления ресурсами здравоохранения*, Краснодар: КубГАУ.
7. Дунская Л.К., Любченко А.А. (2019) Особенности проведения валидации модели при построении прогноза с использованием клеточно-автоматной теории. *Научное обеспечение агропромышленного комплекса: сборник статей по материалам XII Всероссийской конференции молодых ученых*, 87–88.

8. Яковлев С.С. (2023) Сравнительный анализ интеллектуальных методов анализа данных. *Информационные технологии, энергетика и экономика: труды XX Международной научно-технической конференции студентов и аспирантов*, 2, 98–103.
9. Зеленков Ю.А., Анисичкина Е.А. (2021) Динамика исследований в области интеллектуального анализа данных: тематический анализ публикаций за 20 лет. *Бизнес-информатика*, 151 (1), 30–46. DOI: <https://doi.org/10.17323/2587-814X.2021.1.30.46>
10. Будникова И.К., Плетенева Е.В. (2022) Кластерный анализ как функция интеллектуального анализа данных. *Информационные технологии в строительных, социальных и экономических системах*, 1 (27), 25–28.
11. Чернышева Ю.Г., Хирная А.В., Головкин А.А. (2023) Интеллектуальный анализ данных (Data Mining) и его применение в бизнес-анализе. *Бухгалтерский учет, анализ, аудит и статистика: информационные инструменты достижения целей устойчивого развития экономики: материалы Международной научно-практической конференции, посвященной 70-летию Учетно-экономического факультета*, 212–217.
12. Корсакова С.К., Русина Н.В., Бондарик В.М. (2024) Прогнозирование временного ряда в Python с помощью метода Хольта-Уинтерса. *BIG DATA и анализ высокого уровня: сборник научных статей X Международной научно-практической конференции*, 438–443.
13. Tay F.E.H., Shen L. (2002) Economic and financial prediction using rough sets model. *European Journal of Operational Research*, 141 (3), 641–659. DOI: [https://doi.org/10.1016/S0377-2217\(01\)00259-4](https://doi.org/10.1016/S0377-2217(01)00259-4)
14. Cao L., Yang Q., Yu Ph.S (2021) Data science and AI in FinTech: an overview. *International Journal of Data Science and Analytics*, 12 (2), 81–99. DOI: <https://doi.org/10.1007/s41060-021-00278-w>
15. Midoun M., Belbachir H. (2022) A new process for mining spatial databases: combining spatial data mining and visual data mining. *International Journal of Business Information Systems*, 39 (1), 17–51. DOI: <https://doi.org/10.1504/IJBIS.2022.120366>
16. García-Saiz D., Zorrilla M., Bosque J.L. (2017) A clustering-based knowledge discovery process for data centre infrastructure management. *The Journal of Supercomputing*, 73 (1), 215–226. DOI: <https://doi.org/10.1007/s11227-016-1693-z>
17. Ienco D., Bordogna G. (2018) Fuzzy extensions of the DBScan clustering algorithm. *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, 22 (5), 1719–1730. DOI: <https://doi.org/10.1007/s00500-016-2435-0>
18. Sass J., Thös A.K. (2024) Risk reduction and portfolio optimization using clustering methods, *Econometrics and Statistics*, 32, 1–16. DOI: <https://doi.org/10.1016/j.ecosta.2021.11.010>
19. Дунская Л.К., Попова Е.В. (2024) Интеллектуальный анализ данных: кластеризация временных рядов как способ выявления скрытых паттернов поведения. *Цифровизация экономики: направления, методы, инструменты: сборник материалов VI Всероссийской научно-практической конференции*, 112–113.
20. Тотмянин Н.Р., Бондарчук Д.О. (2025) Сравнительный анализ методов анализа данных: корреляция, регрессия, временные ряды. *Славянский форум*, 1 (47), 33–38.
21. Трубицына О.А. (2024) Использование адаптивных методов прогнозирования в эконометрических исследованиях. *Планирование, проведение и толкование итогов научных исследований: сборник статей Международной научно-практической конференции*, 93–96.
22. Клячкин В.Н. (2024) Многомерный статистический контроль процессов. *Стандарты и качество*, 13, 70–81.
23. Петров А.Н. (2024) Прогнозирование на временных рядах с привлечением машинного обучения. *Научно-технический вестник Поволжья*, 1, 136–139.
24. Hossain Md.Z., Islam Md.Ja., Miah Md.W.R., Rony Ja.H., Begum M. (2021) Develop a dynamic dbscan algorithm for solving initial parameter selection problem of the DBSCAN algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 23 (3), 1602–1610. DOI: <http://doi.org/10.11591/ijeecs.v23.i3.pp1602-1610>
25. Gholizadeh N., Saadatfar H., Hanafi N. (2021) K-DBSCAN: an improved DBSCAN algorithm for BIG DATA. *The Journal of Supercomputing*, 77, 6214–6235. DOI: <https://doi.org/10.1007/s11227-020-03524-3>
26. Ramos Emmendorfer L., de Paula Canuto A.M. (2021) A generalized average linkage criterion for hierarchical agglomerative clustering. *Applied Soft Computing*, 100, art. no. 106990. DOI: <https://doi.org/10.1016/j.asoc.2020.106990>



27. Ackermann M.R., Blömer J., Kuntze D., Sohler Ch. (2014) Analysis of Agglomerative Clustering. *Algorithmica*, 69 (1), 184–215. DOI: <https://doi.org/10.1007/s00453-012-9717-4>
28. Seniwati E., Sidauruk A., Haryoko H., Lukman A. (2023) Clustering Performance Between K-Means and Bisecting K-Means for Students Interest in Senior High School. *Building of Informatics, Technology and Science*, 5 (1), 308–316. DOI: <https://doi.org/10.47065/bits.v5i1.3624>
29. Дядюнов Д.А. (2025) Машинное обучение для риск-менеджмента в банке: возможности и вызовы. *Вестник науки*, 1 (1 (82)), 265–273.
30. Джаррар Ф.Ш. (2024) К вопросу об использовании методов обнаружения аномалий при анализе Bigdata. *Цифровая трансформация социальных и экономических систем: материалы международной научно-практической конференции*, 800–805.
31. Купряшин Е.В., Врабие И.В., Сыресин Д.Е. (2021) Подход к оценке достоверности прогноза дебитов в задачах виртуальной расходомерии. *Интеллектуальный анализ данных в нефтегазовой отрасли: материалы II региональной конференции EAGE в России и странах СНГ*, 24. DOI: <https://doi.org/10.3997/2214-4609.202156032>
32. Шайхетдинова А.Р., Лутфуллин Ю.Р. (2023) Оценка качества прогнозов в экономической статистике. *Генезис и онтология инновационно ориентированной деятельности в условиях цифровизации: сборник статей Всероссийской научно-практической конференции с международным участием*, 158–160.

REFERENCES

1. Nicolis G. (1995) *Introduction to Nonlinear Science*, Cambridge, UK: Cambridge University Press.
2. Jagadev A.K., Routra N.K., Sarangi P.P., Patra S.K., Paikaray B.K. (in press) (2024) Fuzzy inventory modelling: addressing uncertainty in economic order quantity analysis within nonlinear science. *International Journal of Applied Nonlinear Science*. DOI: <https://doi.org/10.1504/ijans.2024.10066705>
3. Kumratova A.M., Popova E.V., Savinskaya D.N., Kurnosova N.S. (2015) Kompleksnaya metodika analiza ekonomicheskikh vremennykh ryadov metodami nelineinoy dinamiki. *Sovremennaya ekonomika: problemy i resheniya*, 8 (68), 35–43. DOI: <https://doi.org/10.17308/meps.2015.8/1292>
4. Savinskaya D.N., Velikanova L.O., Dunskey L.K., Popova M.I. (2020) Predprognoznnyy analiz vremennogo ryada s vyyavleniem trend-sezonnykh komponent metodom Chetverikova s ispol'zovaniem sredstv MS Excel. *Sovremennaya ekonomika: problemy i resheniya*, 10 (130), 18–25. DOI: <https://doi.org/10.17308/meps.2020.10/2445>
5. Kumratova A.M., Popova E.V., Kostenko I.V., Dunskey L.K. (2019) Adaptatsiya lineinogo kletchnogo avtomata dlya resheniya zadach prognozirovaniya na baze prirodno-ekonomicheskikh vremennykh ryadov. *Sovremennaya ekonomika: problemy i resheniya*, 9 (117), 8–17. DOI: <https://doi.org/10.17308/meps.2019.9/2196>
6. Kumratova A.M., Popova E.V., Bidzhiev A.Z. (2014) *Ekonomiko-matematicheskoe modelirovanie riska v zadachakh upravleniya resursami zdravookhraneniya*, Krasnodar: KubGAU.
7. Dunskey L.K., Lyubchenko A.A. (2019) Osobennosti provedeniya validatsii modeli pri postroenii prognoza s ispol'zovaniem kletchno-avtomatnoi teorii. *Nauchnoe obespechenie agropromyshlennogo kompleksa: sbornik statei po materialam XII Vserossiiskoi konferentsii molodykh uchenykh*, 87–88.
8. Yakovlev S.S. (2023) Sravnitel'nyi analiz intellektual'nykh metodov analiza dannykh. *Informatsionnye tekhnologii, energetika i ekonomika: trudy XX Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii studentov i aspirantov*, 2, 98–103.
9. Zelenkov Yu.A., Anisichkina E.A. (2021) Dinamika issledovaniy v oblasti intellektual'nogo analiza dannykh: tematicheskii analiz publikatsii za 20 let. *Biznes-informatika*, 151 (1), 30–46. DOI: <https://doi.org/10.17323/2587-814X.2021.1.30.46>
10. Budnikova I.K., Pleteneva E.V. (2022) Klasternyy analiz kak funktsiya intellektual'nogo analiza dannykh. *Informatsionnye tekhnologii v stroitel'nykh, sotsial'nykh i ekonomicheskikh sistemakh*, 1 (27), 25–28.
11. Chernysheva Yu.G., Khirnaya A.V., Golovko A.A. (2023) Intellektual'nyi analiz dannykh (Data Mining) i ego primeneniye v biznes-analize. *Bukhgalterskii uchet, analiz, audit i statistika: informatsionnye instrumenty dostizheniya tselei ustoichivogo razvitiya ekonomiki: materialy Mezhdunarodnoi nauchno-prakticheskoi konferentsii, posvyashchennoi 70-letiyu Uchetno-ekonomicheskogo fakul'teta*, 212–217.

12. Korsakova S.K., Rusina N.V., Bondarik V.M. (2024) Prognozirovanie vremennogo ryada v Python s pomoshch'yu metoda Khol'ta-Uintersa. *BIG DATA i analiz vysokogo urovnya: sbornik nauchnykh statei X Mezhdunarodnoi nauchno-prakticheskoi konferentsii*, 438–443.
13. Tay F.E.H., Shen L. (2002) Economic and financial prediction using rough sets model. *European Journal of Operational Research*, 141 (3), 641–659. DOI: [https://doi.org/10.1016/S0377-2217\(01\)00259-4](https://doi.org/10.1016/S0377-2217(01)00259-4)
14. Cao L., Yang Q., Yu Ph.S (2021) Data science and AI in FinTech: an overview. *International Journal of Data Science and Analytics*, 12 (2), 81–99. DOI: <https://doi.org/10.1007/s41060-021-00278-w>
15. Midoun M., Belbachir H. (2022) A new process for mining spatial databases: combining spatial data mining and visual data mining. *International Journal of Business Information Systems*, 39 (1), 17–51. DOI: <https://doi.org/10.1504/IJBIS.2022.120366>
16. García-Saiz D., Zorrilla M., Bosque J.L. (2017) A clustering-based knowledge discovery process for data centre infrastructure management. *The Journal of Supercomputing*, 73 (1), 215–226. DOI: <https://doi.org/10.1007/s11227-016-1693-z>
17. Ienco D., Bordogna G. (2018) Fuzzy extensions of the DBScan clustering algorithm. *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, 22 (5), 1719–1730. DOI: <https://doi.org/10.1007/s00500-016-2435-0>
18. Sass J., Thös A.K. (2024) Risk reduction and portfolio optimization using clustering methods, *Econometrics and Statistics*, 32, 1–16. DOI: <https://doi.org/10.1016/j.ecosta.2021.11.010>
19. Dunsakaya L.K., Popova E.V. (2024) Intel'ektual'nyi analiz dannykh: klasterizatsiya vremennykh ryadov kak sposob vyavleniya skrytykh patternov povedeniya. *Tsifrovizatsiya ekonomiki: napravleniya, metody, instrumenty: sbornik materialov VI Vserossiiskoi nauchno-prakticheskoi konferentsii*, 112–113.
20. Totmyanin N.R., Bondarchuk D.O. (2025) Sravnitel'nyi analiz metodov analiza dannykh: korrelyatsiya, regressiya, vremennyye ryady. *Slavyanskii forum*, 1 (47), 33–38.
21. Trubitsyna O.A. (2024) Ispol'zovanie adaptivnykh metodov prognozirovaniya v ekonometricheskikh issledovaniyakh. *Planirovanie, provedenie i tolkovanie itogov nauchnykh issledovaniy: sbornik statei Mezhdunarodnoi nauchno-prakticheskoi konferentsii*, 93–96.
22. Klyachkin V.N. (2024) Mnogomernyi statisticheskii kontrol' protsessov. *Standarty i kachestvo*, 13, 70–81.
23. Petrov A.N. (2024) Prognozirovanie na vremennykh ryadakh s privilecheniem mashinnogo obucheniya. *Nauchno-tekhnicheskii vestnik Povolzh'ya*, 1, 136–139.
24. Hossain Md.Z., Islam Md.Ja., Miah Md.W.R., Rony Ja.H., Begum M. (2021) Develop a dynamic dbscan algorithm for solving initial parameter selection problem of the DBSCAN algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 23 (3), 1602–1610. DOI: <http://doi.org/10.11591/ijeecs.v23.i3.pp1602-1610>
25. Gholizadeh N., Saadatfar H., Hanafi N. (2021) K-DBSCAN: an improved DBSCAN algorithm for BIG DATA. *The Journal of Supercomputing*, 77, 6214–6235. DOI: <https://doi.org/10.1007/s11227-020-03524-3>
26. Ramos Emmendorfer L., de Paula Canuto A.M. (2021) A generalized average linkage criterion for hierarchical agglomerative clustering. *Applied Soft Computing*, 100, art. no. 106990. DOI: <https://doi.org/10.1016/j.asoc.2020.106990>
27. Ackermann M.R., Blömer J., Kuntze D., Sohler Ch. (2014) Analysis of Agglomerative Clustering. *Algorithmica*, 69 (1), 184–215. DOI: <https://doi.org/10.1007/s00453-012-9717-4>
28. Seniwati E., Sidauruk A., Haryoko H., Lukman A. (2023) Clustering Performance Between K-Means and Bisecting K-Means for Students Interest in Senior High School. *Building of Informatics, Technology and Science*, 5 (1), 308–316. DOI: <https://doi.org/10.47065/bits.v5i1.3624>
29. Dyadyunov D.A. (2025) Mashinnoe obuchenie dlya risk-menedzhmenta v banke: vozmozhnosti i vyzovy. *Vestnik nauki*, 1 (1 (82)), 265–273.
30. Dzharrar F.Sh. (2024) K voprosu ob ispol'zovanii metodov obnaruzheniya anomalii pri analize Bigdata. *Tsifrovaya transformatsiya sotsial'nykh i ekonomicheskikh sistem: materialy mezhdunarodnoi nauchno-prakticheskoi konferentsii*, 800–805.
31. Kupryashin E.V., Vrabie I.V., Syresin D.E. (2021) Podkhod k otsenke dostovernosti prognoza debitov v zadachakh virtual'noi raskhodometrii. *Intel'ektual'nyi analiz dannykh v neftegazovoi otrasli: materialy II regional'noi konferentsii EAGE v Rossii i stranakh SNG*, 24. DOI: <https://doi.org/10.3997/2214-4609.202156032>



32. Shaikhetdinova A.R., Lutfullin Yu.R. (2023) Otsenka kachestva prognozov v ekonomicheskoi statistike. *Genezis i ontologiya innovatsionno orientirovannoi deyatel'nosti v usloviyakh tsifrovizatsii: sbornik statei Vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem*, 158–160.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

ДУНСКАЯ Лада Константиновна

E-mail: dunskaia.l@edu.kubsau.ru

Lada K. DUNSKAIA

E-mail: dunskaia.l@edu.kubsau.ru

ORCID: <https://orcid.org/0009-0008-1844-2620>

ПОПОВА Елена Витальевна

E-mail: popova.e@kubsau.ru

Elena V. POPOVA

E-mail: popova.e@kubsau.ru

ORCID: <https://orcid.org/0000-0001-8533-6897>

Поступила: 09.01.2025; Одобрена: 15.02.2025; Принята: 16.02.2025.

Submitted: 09.01.2025; Approved: 15.02.2025; Accepted: 16.02.2025.