

ПРИМЕНЕНИЕ КАРТЫ КОХОНЕНА ДЛЯ АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ РЕГИОНОВ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ИХ ОПТИМИЗАЦИИ

Д. Б. Пугачева, А. В. Юдин

Пугачева Дарья Борисовна, студент, МИРЭА – Российский технологический университет, г. Москва, Российская Федерация, тел.: + 7 (910) 773-92-25; E-mail: d.pugacheva@list.ru;

Юдин Александр Викторович, доктор экономических наук, кандидат физико-математических наук, доцент, и. о. заведующего кафедрой индустриального программирования, Институт перспективных технологий и индустриального программирования МИРЭА – Российского технологического университета, г. Москва, Российская Федерация, тел.: + 7 (965) 343-25-61; e-mail: yudinmsk@yandex.ru

В данной статье предлагается решение для анализа и оптимизации данных с использованием самоорганизующейся карты Кохонена. Разработан модифицированный алгоритм для формирования и дальнейшего преобразования карты на основе цветных коэффициентов, соотнесенных с исходными данными. Предлагаемое решение отличается высокой наглядностью представления как исходных данных, так и полученных после преобразования результатов, а также простотой последующей интерпретации. Алгоритм, реализованный на языке программирования Python, предусматривает использование данных без предварительной обработки, данных обработанных и приведенных в формат цветных коэффициентов от 0 до 255 (формат Red Green Blue), шестнадцатеричный формат (HEX), а также данных, преобразованных в цветные коэффициенты, используемые библиотекой Matplotlib. Рассмотрено применение нейронной сети для анализа входных данных в формате одномерного массива с двумя фиксированными цветовыми коэффициентами и трех одномерных массивов, где каждый массив исходных данных сопоставляется с одним из трех цветных коэффициентов. Предложенный метод рассматривается в отношении данных Федеральной службы государственной статистики, а именно: Рентабельность продаваемых товаров, продукции (работ, услуг) организаций по добыче полезных ископаемых, обрабатывающих производств, по обеспечению электрической энергией, газом и паром; кондиционированию воздуха в 2021 году по данным бухгалтерской отчетности с использованием Государственного информационного ресурса бухгалтерской отчетности ФСН России. Полученные результаты могут стать основой для дальнейших исследований в области анализа данных и принятия решений на распределение ресурсов субъектов РФ и последующего прогнозирования уровня развития регионов во времени.

Ключевые слова: *нейронная сеть, Python, TensorFlow, карта Кохонена, оптимизация данных, обработка данных, развитие регионов.*

APPLICATION OF THE KOHONEN MAP FOR THE ANALYSIS OF STATISTICAL DATA OF REGIONS TO SUPPORT DECISION-MAKING ON THEIR OPTIMIZATION

D. B. Pugacheva, A. V. Yudin

Pugacheva Dariya Borisovna, student, MIREA – Russian University of Technology, Moscow, Russian Federation, phone: + 7 (910) 773-92-25; e-mail: d.pugacheva@list.ru;

Yudin Aleksandr Viktorovich, Doctor of Economic Sciences, Candidate of Physical and Mathematical Sciences, Associate Professor, Acting Head of the Department of Industrial Programming, Institute of Advanced Technologies and Industrial Programming of MIREA – Russian Technological University, Moscow, Russian Federation, phone: + 7 (965) 343-25-61; e-mail: yudinmsk@yandex.ru

This article proposes a solution for data analysis and optimization using a self-organizing Kohonen map. A modified algorithm has been developed for the formation and further transformation of the map based on color coefficients correlated with the original data. The proposed solution is characterized by high visibility of both the initial data and the results obtained after conversion, as well as the simplicity of subsequent interpretation. The algorithm, implemented in the Python programming language, provides for the use of data without preliminary processing, data processed and converted into color coefficients from 0 to 255 (Red Green Blue format), hexadecimal format (HEX), as well as data converted into color coefficients used by the Matplotlib library. The use of a neural network for analyzing input data in the format of a one-dimensional array with two fixed color coefficients and three one-dimensional arrays, where each array of source data is compared with one of three color coefficients, is considered. The proposed method is considered in relation to the data of the Federal State Statistics Service, namely: The profitability of goods sold, products (works, services) of mining organizations, manufacturing industries, providing electric energy, gas and steam; air conditioning in 2021 according to accounting reports using the State Information Resource of accounting reports of the Federal Tax Service of Russia. The obtained results can become the basis for further research in the field of data analysis and decision-making on the allocation of resources of the subjects of the Russian Federation and subsequent forecasting of the level of development of regions over time.

Keywords: *neural network, Python, TensorFlow, Kohonen map, data optimization, data processing, regional development.*

В настоящей работе решается задача представления наглядного решения для анализа и оптимизации статистических показателей регионов Российской Федерации посредством создания карты Кохонена. Предлагается решение, обеспечивающее наглядность и информативность исследуемых и преобразованных данных. Оптимизация данных с использованием нейронной сети позволит обеспечить соседние узлы схожими уровнями показателей, минимизировать разрывы между избыточными и недостаточными, отрицательными значениями. Полученные данные могут быть применимы для принятия решения о распределении ресурсов по регионам.

Анализ достижений и публикаций, в которых предлагается решение данной проблемы или задачи

Схожая модель применения самоорганизующихся карт Кохонена была описана ранее: в статьях [1] и [2]. Однако, данные исследования направлены на кластеризацию и выявление паттернов. Также стоит выделить статью «Оптимизация обработки данных на основе модифицированного обучения нейронной сети и сегментации случайных временных процессов» [3], где рассматривается задача оптимизации обработки данных.

Так, работа [1] посвящена вопросам анализа инновационного развития производства в регионах Российской Федерации с применением карт Кохонена. В качестве исследуемого показателя, определяющего инновационное развитие производства в регионах РФ, использован показатель – число разработанных передовых производственных технологий за период 2010–2014 годы по всем видам производств для 72 субъектов РФ. С применением карт Кохонена выделено 6 кластеров.

В работе [2] на основе подхода системной экономической теории предпринята попытка формирования альтернативной методики мониторинга результатов государственного управления. В ходе исследования осуществляется кластеризация субъектов РФ в зависимости от уровня социально-экономического развития в пространственно-временном разрезе.

В работе [3] разработаны и представлены методы оптимизации, идентификации и обработки данных на основе радиально-базисной сети.

Исследовательская часть

Использование самоорганизующихся карт Кохонена в подобных исследованиях находит применение в следующих областях анализа данных:

1) кластеризация и выявление паттернов: карта Кохонена способна к кластеризации данных, выделяя схожие регионы или факторы, влияющие на устойчивое развитие, что позволяет выявить общие паттерны и тенденции. Именно данная область анализа данных является основной в вышеперечисленных исследованиях;

2) идентификация приоритетных областей: анализ кластеров на карте может помочь идентифицировать регионы, которые требуют особого внимания в плане устойчивого развития, что может стать отправной точкой для целенаправленных усилий и инвестиций;

3) прогнозирование и стратегическое планирование: на основе выявленных паттернов и кластеров можно разрабатывать прогнозы и стратегии для устойчивого развития, что обеспечивает более обоснованный подход к принятию решений;

4) оптимизация ресурсов: анализ карты Кохонена может помочь определить, как распределены ресурсы в различных областях и где возможны оптимизации для повышения эффективности устойчивого развития;

5) отслеживание изменений во времени: обновление карты в соответствии с публикуемыми открытыми данными позволяет отслеживать изменения в развитии регионов с течением времени и адаптировать стратегии соответственно.

Таким образом, использование самоорганизующейся карты Кохонена предоставляет инструмент для системного анализа данных, что может улучшить понимание факторов, влияющих на устойчивое развитие страны, развитие в целом.

В данной работе предложено и рассмотрено использование карты Кохонена для анализа данных и применения полученных результатов для оптимизации распределения ресурсов.

Преимущества использования нейронной сети Кохонена включают выдающуюся способность к кластеризации данных, обучение без учителя, а также эффективное отображение сложных многомерных пространств, что делает ее мощным

инструментом для анализа и обработки информации в различных областях, от распознавания образов до оптимизации процессов принятия решений.

В качестве информационной основы примем данные Росстата [4] о рентабельности проданных товаров, продукции (работ, услуг) организаций по добыче полезных ископаемых, обрабатывающих производств, по обеспечению электронной энергией, газом и паром; кондиционированию воздуха в 2021 году [5].

В ходе работы была разработана нейронная сеть Кохонена и оптимизирована Самоорганизующаяся карта Кохонена для возможности анализа данных и принятия решений на управление, оптимизацию.

Для реализации программного продукта, обучения и использования самоорганизующейся нейронной сети Кохонена был выбран язык программирования Python и библиотека TensorFlow, что обосновано несколькими ключевыми причинами. Во-первых, Python – это широко используемый язык программирования с простым и читаемым синтаксисом, что облегчает разработку и поддержку кода. Благодаря обширной экосистеме библиотек и фреймворков, Python стал основным выбором в области машинного обучения и искусственного интеллекта.

TensorFlow, в свою очередь, предоставляет мощный набор инструментов для построения и обучения нейронных сетей. Его гибкость и высокая производительность делают его оптимальным для реализации и оптимизации нейронных сетей, включая сети Кохонена. Библиотека предоставляет простой интерфейс для создания сложных моделей, а также оптимизированные инструменты для эффективного использования аппаратных ресурсов.

Таким образом, выбор Python и TensorFlow обеспечивает эффективное и удобное окружение для разработки, обучения и применения нейронных сетей Кохонена в контексте анализа данных.

Дополнительными, используемыми в ходе реализации, библиотеками являлись NumPy и Matplotlib, используемые для работы с многомерными массивами и визуализации данных соответственно.

Первым этапом после реализации являлось обучение обучения самоорганизующейся нейронной сети и определения весов нейронов. Далее была проведена предварительная обработка данных. Исходными данными для анализа были выбраны Рентабельность продаваемых товаров, продукции (работ, услуг) организаций по добыче полезных ископаемых, обрабатывающих производств, по обеспечению электрической энергией, газом и паром; кондиционированию воздуха в 2021 году по данным бухгалтерской отчетности с использованием Государственного информационного ресурса бухгалтерской отчетности ФНС России (открытые данные Федеральной службы государственной статистики). Выбор этих данных (представлены на рисунке 1) был обоснован необходимостью использования двумерного массива данных [3; n], что обеспечит возможность преобразования и связи каждого элемента массива с цветовой информацией (коэффициенты Red Green Blue) для соответствующего элемента карты Кохонена. Использование данной связи обеспечивает ассоциацию цветов с определенными признаками, классами и данными.

13.13. РЕНТАБЕЛЬНОСТЬ ПРОДАННЫХ ТОВАРОВ, ПРОДУКЦИИ (РАБОТ, УСЛУГ) ОРГАНИЗАЦИЙ ПО ДОБЫЧЕ ПОЛЕЗНЫХ ИСКОПАЕМЫХ, ОБРАБАТЫВАЮЩИХ ПРОИЗВОДСТВ, ПО ОБЕСПЕЧЕНИЮ ЭЛЕКТРИЧЕСКОЙ ЭНЕРГИЕЙ, ГАЗОМ И ПАРОМ; КОНДИЦИОНИРОВАНИЮ ВОЗДУХА в 2021 г.¹⁾
 (в процентах)

	Добыча полезных ископаемых	Обрабатывающие производства	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха		Добыча полезных ископаемых	Обрабатывающие производства	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха
Российская Федерация	35,5	16,2	9,4				
Центральный федеральный округ	47,9	13,7	13,2	Приволжский федеральный округ	23,9	16,3	5,7
Белгородская область	287,8	8,6	2,8	Республика			
Брянская область	-28,9	9,7	2,9	Башкортостан	15,9	15,6	6,8
Владимирская область	1,2	24,8	4,3	Республика Марий Эл	25,3	8,2	5,6
Воронежская область	12,4	13,9	2,5	Республика Мордовия	-	16,9	2,1
Ивановская область	-3,7	9,5	1,7	Республика Татарстан	23,8	13,7	5,3
Калужская область	4,4	6,4	2,4	Удмуртская Республика	19,8	9,3	2,9
Костромская область	13,0	12,9	2,0	Чувашская Республика	-6,8	10,4	4,2
Курская область	177,4	8,0	7,2	Пермский край	29,5	42,0	3,2
Липецкая область	5,6	28,0	3,2	Кировская область	-0,2	13,9	1,0
Московская область	4,5	11,0	7,8	Нижегородская область	15,3	14,5	8,6
Орловская область	2,8	20,8	2,9	Оренбургская область	23,4	13,2	41,0
Рязанская область	-8,0	12,5	4,5	Пензенская область	8,2	13,0	5,3
Смоленская область	-26,4	17,4	4,5	Самарская область	24,8	16,8	4,9
Тамбовская область	-	11,6	3,0	Саратовская область	22,2	12,6	1,4
Тверская область	54,7	6,2	-5,6	Ульяновская область	23,7	5,2	1,6
Тульская область	11,7	17,1	7,1	Уральский федеральный округ	28,5	21,3	7,0
Ярославская область	11,1	12,0	4,9	Курганская область	3,6	8,0	4,2
г. Москва	17,4	14,1	17,0	Свердловская область	125,8	13,8	4,2
Северо-Западный федеральный округ	41,2	16,0	5,1	Тюменская область	27,5	34,7	10,7
Республика Карелия	197,4	23,4	-2,5	в том числе:			
Республика Коми	12,5	31,7	3,0	Ханты-Мансийский автономный округ – Югра			11,2
Архангельская область	36,2	19,3	-0,7	Ямало-Ненецкий автономный округ	21,2	19,7	4,3
в том числе:				Тюменская область без автономных округов	30,9	42,8	14,2
Ненецкий автономный округ	17,0	-14,2	-6,1	Челябинская область	44,6	22,2	3,6
Архангельская область без автономных округов	97,2	19,4	0,1	Сибирский федеральный округ	45,5	29,6	18,0
Вологодская область	0,8	44,4	3,4	Республика Алтай	35,8	8,5	-2,3
Калининградская область	31,8	-3,2	7,5	Республика Тыва	85,1	6,1	1,7
Ленинградская область	0,5	15,7	8,4	Республика Хакасия	47,5	7,7	39,6
Мурманская область	126,1	50,0	-5,2	Алтайский край	14,1	26,2	6,2
Новгородская область	-8,7	55,4	0,0	Красноярский край	58,6	58,5	29,1
Псковская область	11,6	5,2	0,7	Иркутская область	39,3	20,8	27,6
г. Санкт-Петербург	77,6	8,2	6,8	Кемеровская область	48,8	15,3	7,8
Южный федеральный округ	58,7	8,8	4,2	Новосибирская область	53,8	11,0	3,7
Республика Адыгея	13,2	13,2	0,0	Омская область	-67,5	15,4	7,7
Республика Калмыкия	-26,4	1,1	60,2	Томская область	10,9	13,5	3,9
Республика Крым	-3,6	12,3	-7,6	Дальневосточный федеральный округ	62,0	9,7	-5,4
Краснодарский край	8,7	10,4	5,4	Республика Бурятия	28,6	28,0	-12,8
Астраханская область	80,9	-14,6	6,8	Республика Саха (Якутия)	72,1	0,1	-0,4
Волгоградская область	12,7	9,9	2,7	Забайкальский край	105,8	-8,8	-0,4
Ростовская область	27,3	7,0	6,0	Камчатский край	58,9	34,8	4,7
г. Севастополь	25,0	3,4	-4,0	Приморский край	6,3	-0,4	-5,8
Северо-Кавказский федеральный округ	8,0	23,6	1,6	Хабаровский край	80,1	13,9	-9,2
Республика Дагестан	-5,4	19,8	-27,0	Амурская область	32,5	5,1	-0,9
Республика Ингушетия	12,4	5,9	-46,8	Магаданская область	60,9	26,1	-10,1
Кабардино-Балкарская Республика	-100,0	5,6	-16,7	Сахалинская область	19,4	38,6	-14,5
Карачаево-Черкесская Республика	7,1	8,0	-3,9	Еврейская автономная область	81,1	-1,1	-22,3
Республика Северная Осетия – Алания	8,4	10,9	-18,4	Чукотский автономный округ	95,1	-8,9	-25,8
Чеченская Республика	4,5	2,1	-20,9				
Ставропольский край	11,9	29,3	9,2				

¹⁾ По данным бухгалтерской отчетности с использованием Государственного информационного ресурса бухгалтерской отчетности ФНС России. Знак (-) означает убыточность.

Рис. 1. Используемые данные

Для корректной работы нейронной сети Кохонена была проведена нормировка данных путем соотношения и представления их числовым значениям цвета RGB. После описанного преобразования данные принимают значения от 0 до 256 пропорционально исходным (с учетом убыточных показателей). Так, наибольшие значения на исходной карте будут представлены явно выделенными красным, зеленым и синим цветами (при использовании для визуализации всех трех столбцов данных, что описано во втором этапе тестирования). Напротив, наименьшие или убыточные значения будут иметь сине-зеленый, фиолетовый и красно-зеленый оттенки.

Следующим этапом работы являлось тестирование самоорганизующейся карты Кохонена для анализа массива, состоящего из исходных данных и двух фиксированных параметров. Для корректного отображения цвета и наглядности результатов использованы данные в следующем формате:

• коэффициент Red – преобразованное в коэффициент значение столбца Обеспечение

электрической энергией, газом и паром; кондиционирование воздуха;

• коэффициент Green – фиксирован на значении 0,5 (128);

• Коэффициент Blue – фиксирован на значении 0.

Значение второго и третьего коэффициента являются фиксированными для корректного отображения карты Кохонена.

Стоит отметить, что показатели, представленные в таблице, были также предварительно отсортированы и отображаются на карте Кохонена в порядке кодов регионов России. Ввиду того, что исходные данные датируются 2021 годом, все новые регионы и регионы, по которым отсутствуют значения и показатели, были представлены на карте с кодом 99. Результат работы представлен на рисунке 2.

При проведении второго этапа тестирования анализировался массив, состоящий исключительно из исходных данных:

• коэффициент Red – преобразованное в коэффициент значение столбца «Добыча полезных ископаемых»;

- коэффициент Green – преобразованное в коэффициент значение столбца «Обрабатывающие производства»;

- коэффициент Blue – преобразованное в коэффициент значение столбца «Обеспечение электрической энергией, газом и паром; кондиционирование воздуха».

Результат работы самоорганизующейся карты Кохонена для описанных данных представлен на рисунке 3.

На исходной карте также наглядно отображаются максимальные значения по рентабельности организаций в области добычи полезных ископаемых – 31 регион (Белгородская область) с максимальным значением коэффициента красного; по рентабельности предприятий, относящихся к обрабатывающим производствам – 53 регион (Новгородская область) с максимальным значением коэффициента зеленого; по обеспечению электрической энергией, газом и

паром; кондиционирование воздуха – 8 регион (Республика Калмыкия).

С минимальными показателями соответственно отображаются 7 регион (Кабардино-Балкарская Республика) с отсутствием коэффициента красного, 30 регион (Астраханская область) с отсутствием коэффициента зеленого и 6 регион (Республика Ингушетия) с отсутствием коэффициента синего.

Полученные результаты представляются наглядно, также предусмотрен вывод полученных цветовых коэффициентов для последующей интерпретации. Исходя из исходных данных формируется оптимальное распределение ресурсов между регионами с избыточным и убыточным производствами. Безусловно, ввиду отсутствия данных по отдельным и новым регионам страны, достоверность полученных результатов не является оптимальной.

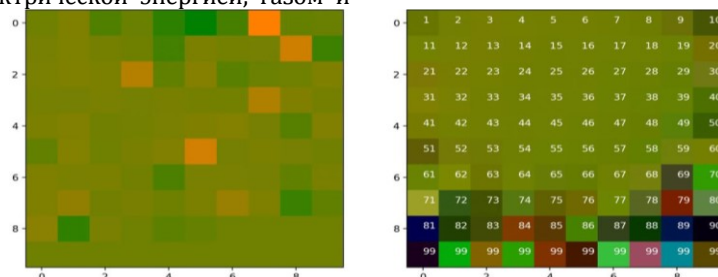


Рис. 2. Результаты работы нейронной сети для двух фиксированных параметров и столбца исходных данных

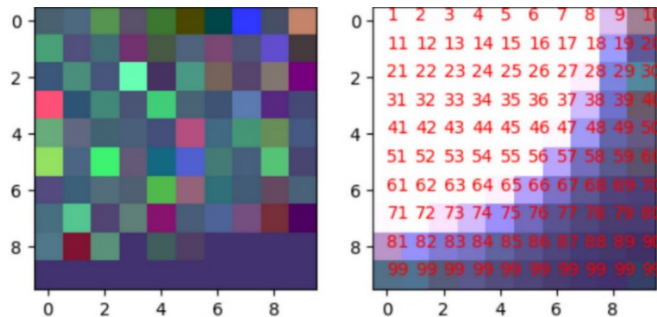


Рис. 3. Результаты работы нейронной сети для трех столбцов исходных данных

Сравнительные диаграммы исходных данных и данных, полученных после преобразования нейронной сетью представлены на рисунках 4–6.

Предлагаемый алгоритм

Применение нейронной сети Кохонена позволяет достаточно эффективно визуализировать медианные и критические показатели при рассмотрении одномерного и двумерного $[3; n]$ массива исходных данных.

Полученные данные могут быть применимы для принятия решения на распределение избыточных или недостаточных ресурсов по регионам.

Для применения описанного решения предлагается следующий алгоритм:

1. Определение и подготовка исходных данных. Возможно использование одномерного массива или массива размерностью $[3; n]$. Оптимальным считается набор данных, где число элементов массива или строк (в случае двумерного массива) является полным квадратом.

2. При использовании одномерного массива

данных проводится предварительное преобразование в массив размерностью $[3; n]$, где первый столбец – полученные коэффициенты от 0 до 255, второй столбец – среднее фиксированное значение, третий столбец – обнуленный коэффициент. При использовании двумерного массива исходных данных каждый столбец пропорционально преобразуется в коэффициенты от 0 до 255.

3. Применение обученной нейронной сети к преобразованным данным.

4. Выгрузка и интерпретация полученных данных, где для большей наглядности рекомендуется использовать веса каждого значения, преобразованного из полученного коэффициента,

Положительные стороны предлагаемого решения: наглядность, простота интерпретации полученных данных. Недостатком является наличие погрешностей при переводе данных в цветовые коэффициенты, а также сложность обучения сети ввиду ограниченного количества представляемых данных рассматриваемой направленности.



Рис. 4. Добыча полезных ископаемых



Рис. 5. Обрабатывающие производства



Рис. 6. Обеспечение электрической энергией, газом и паром; кондиционирование воздуха

Дальнейшее развитие и использование возможно при оптимизации алгоритма для анализа любых входных данных по тем или иным

показателям. Обучение сети на большем количестве схожих данных также позволит увеличить достоверность и качество итогового результата.

Список литературы

1. Куликова О.М., Калугин В.Е., Пильник Н.Б., Гущина А.А. Исследование инновационного развития производства в регионах РФ с применением карт Кохонена // *Фундаментальные исследования*. – 2015. – № 2-25. – С. 5639-5643; URL: <https://fundamental-research.ru/ru/article/view?id=38479> (дата обращения: 23.11.2023).
2. Гагарина Г. Ю., Губарев Р. В., Дзюба Е. И., Файзуллин Ф. С. Прогнозирование социально-экономического развития российских регионов // *Экономика региона*. — 2017. — Т. 13, вып. 4. — С. 1080-1094
3. Холмонов Сунатилло Махмудович Оптимизация обработки данных на основе модифицированного обучения нейронной сети и сегментации случайных временных процессов // *Проблемы информатики*. 2017. №1 (34). URL: <https://cyberleninka.ru/article/n/optimizatsiya-obrabotki-dannyh-na-osnove-modifitsirovannogo-obucheniya-neyronnoy-seti-i-segmentatsii-sluchaynyh-vremennyh-protsessov> (дата обращения: 06.12.2023).
4. Официальная статистика [Электронный ресурс] // Федеральная служба государственной статистики URL: <https://rosstat.gov.ru> (Дата обращения: 28.10.2023).
5. Регионы России. Социально-экономические показатели. 2022: Р32 Стат. сб. / Росстат. – М., 2022. – 1122 с. URL: https://rosstat.gov.ru/storage/mediabank/Region_Pokaz_2022.pdf (дата обращения: 28.10.2023).

© Д. Б. Пугачева, А. В. Юдин

Ссылка для цитирования:

Пугачева Д. Б., Юдин А. В. Применение карты Кохонена для анализа статистических данных регионов для поддержки принятия решений по их оптимизации // *Инженерно-строительный вестник Прикаспия: научно-технический журнал / Астраханский государственный архитектурно-строительный университет*. Астрахань: ГБОУ АО ВО «АГАСУ», 2024. № 1 (47). С. 145–150.