

## РАЗРАБОТКА ПОДХОДА К КЛАСТЕРИЗАЦИИ РАЙОНОВ НА БАЗЕ ИНСТРУМЕНТАЛЬНОГО СРЕДСТВА МАШИННОГО ОБУЧЕНИЯ KNIME

Е.М. Смирнова, А.А. Валинурова, С.В. Данилова, Т.Р. Валинуров

Екатерина Михайловна Смирнова\* (ORCID 0000-0001-7104-6063), Анна Александровна Валинурова (ORCID 0000-0003-4614-7879), Светлана Вадимовна Данилова (ORCID 0000-0002-7879-7713), Тимур Раисович Валинуров (ORCID 0000-0001-9620-0338)

Ивановский государственный университет, ул. Ермака, 37, Иваново, 153025, Россия

E-mail: smirem1@yandex.ru\*, avalinurova@mail.ru, swdanilova@mail.ru, valinurovtr@mail.ru

*Научная работа посвящена применению методов кластеризации для группировки однородных районов в рамках более крупных территориальных образований с целью организации обоснованного государственного финансирования развития территорий. В настоящее время для объединения районов в группы используется ряд статистических показателей, которые не всегда отражают реальную ситуацию и не позволяют обоснованно сгруппировать разнородные районы внутри одной области. Кластеризация для целей группировки может стать инструментом бюджетной политики, позволяющей обоснованно распределить средства, основываясь не только на социальных и демографических показателях, но и опираясь на ряд дополнительных факторов. Районы, имеющие схожие показатели социально-экономического развития, могут иметь разный потенциал и нуждаться в разных стратегиях финансирования. Для принятия обоснованных долгосрочных решений применение метода кластеризации должно основываться на ряде статистических показателей за длительные сроки. В работе приведен подход к кластеризации районов на примере Ивановской области. Областной центр выделен в отдельный кластер, что обосновано существенными отличиями от других населённых пунктов по всем показателям. Кластеризация представляет собой метод машинного обучения без учителя, то есть данные будут группироваться без указания аналитиком критериев разделения.*

**Ключевые слова:** интеллектуальный анализ данных, классификация, кластеризация, район, однородность, социально-экономическое положение, территориальное образование.

## DEVELOPMENT OF AN APPROACH TO CLUSTERING OF DISTRICTS BASED ON THE KNIME MACHINE LEARNING TOOL

E.M. Smirnova, A.A. Valinurova, S.V. Danilova, T.R. Valinurov

Ekaterina M. Smirnova\* (ORCID 0000-0001-7104-6063), Anna A. Valinurova (ORCID 0000-0003-4614-7879), Svetlana V. Danilova (ORCID 0000-0002-7879-7713), Timur R. Valinurov (ORCID 0000-0001-9620-0338)

Ivanovo State University, Ermaka St., 37, Ivanovo, 153025, Russia

E-mail: smirem1@yandex.ru\*, avalinurova@mail.ru, swdanilova@mail.ru, valinurovtr@mail.ru

*The scientific work is devoted to the application of clustering methods for grouping homogeneous areas within larger territorial formations to organize reasonable state financing of territorial development. At the present time for the grouping of districts a number of statistical indicators are used which do not always reflect the real situation and do not allow to reasonably group heterogeneous districts within one area. Clustering for clustering purposes can become a tool of budget policy that allows to reasonably allocate funds based not only on social and demographic indicators, but also on a number of additional factors. Districts with similar indicators of socio-economic development may have different potential and need different financing strategies. To make informed long-term decisions the application of clustering method should be based on a number of statistical indicators over a long period of time. This paper presents an approach to the clustering of districts by the example of Ivanovo region. The regional center is singled out into a separate cluster, which is substantiated by significant differences*

*from other settlements according to all indicators. Clustering is a method of machine learning without a teacher, that is, the data will be grouped without the analyst specifying the criteria of division.*

**Keywords:** intelligent data analysis, classification, clustering, district, homogeneity, socio-economic status, territorial entity.

**Для цитирования:**

Смирнова Е.М., Валинурова А.А., Данилова С.В., Валинуров Т.Р. Разработка подхода к кластеризации районов на базе инструментального средства машинного обучения KNIME. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2021. № 04(50). С.165-175. DOI: 10.6060/ivecofin.2021504.579

**For citation:**

Smirnova E.M., Valinurova A.A., Danilova S.V., Valinurov T.R. Development of an approach to clustering of districts based on the KNIME machine learning tool. *Ivecofin*. 2021. № 04(50). С.165-175. DOI: 10.6060/ivecofin.2021504.579 (in Russian)

## ВВЕДЕНИЕ

Комплексное исследование возможностей использования технологии интеллектуального анализа данных (Data Mining) становится актуальным для решения наиболее сложных проблем социально-экономического характера, выработки стратегии и принятия управленческих решений по развитию отдельно взятого района территориального образования, использования данных для разработки региональной программы развития и решение задач распределения государственной поддержки [1, 12]. Применение указанной технологии позволит произвести разработку и первичную настройку концептуальной модели поиска новых, ранее неизвестных и практически полезных закономерностей развития районов Ивановской области с целью оценки социально-экономического положения территориального образования и выработки практических рекомендаций по повышению инвестиционной привлекательности региона. Отметим, что данное направление развития соответствует стратегии развития новых цифровых технологий в РФ [13].

## МАТЕРИАЛЫ И МЕТОДЫ

В настоящее время учеными-исследователями, региональными органами власти и аналитическими агентствами используется метод классификации районов по отдельным показателям. Его применение носит субъективный, выборочный и непоследовательный характер, отнесение к той или иной группе производится на основании анализа двух групп показателей (социально-экономическое развитие и демографическое положение районов), что подвергается критике, поскольку не учитывает множество дополнительных факторов. Существующая система классификации обладает рядом существенных недостатков и требует пересмотра. Чтобы повысить объективность исследования предлагается использовать кластеризацию

с помощью интеллектуального анализа данных для выявления новые закономерности в группировке районов Ивановской области [6, 18].

В работе также будут использованы такие методы как: анализ и синтез данных, метод группировок, метод графического отражения данных.

## РЕЗУЛЬТАТЫ

Основным научным результатом работы можно считать новый подход к разделению районов Ивановской области по степени однородности, проведенный посредством метода кластеризации. Предложенный подход рекомендуется использовать для выработки политики в области государственного финансирования.

## ДИСКУССИЯ

По мнению Б.Г. Миркина под кластером обычно понимается часть данных, выделяемая из остальной части наличием некоторой однородности ее элементов, которыми является подмножество объектов или подмножество переменных [14]. Таким образом, кластерный анализ или кластеризацию с появлением машинного обучения начали определять как способ группировки многомерных объектов (имеющих несколько независимых входных величин), базирующихся на представлении результатов каких-либо измерений в графическом виде с использованием точек наиболее подходящего геометрического пространства. Путем группировки этих точек и выделения их скопления происходит формирование кластеров [2, 8, 11]. Кластеризацию относится к ключевым методам Data Mining, при котором осуществляется поиск заранее неизвестных ответов, т.е. система самостоятельно находит корреляцию данных и происходит обучение без учителя.

Кластеризация может быть использована для выделения районов по схожим признакам, бу-

дет базой для создания групп однотипных районов для применения к ним одинаковых инструментов воздействия в рамках их социально-экономического развития. Исходные данные для исследования – паспорта муниципальных образований в формате Excel с информацией о социально-экономических показателях за 2020 г.

Для регионов, которые ставят перед собой задачу определения приоритетных и комплексных направлений социально-экономического развития муниципальных образований, возможно использование результатов кластерного анализа в нескольких областях. Первый аспект применения – это выявление ключевых проблем и формирование перечня тех районов, у которых низкие социально-экономические показатели и которым требуется первоочередная помощь. Второе направление применения кластеризации – это оценка «потенциала территории», отбор тех районов, которые могут стать драйверами роста, под которые в дальнейшем могут создаваться инвестиционные проекты развития, индустриальные парки, особые экономические зоны. Финансирование данных проектов будет полностью или частично за счет бюджетных средств. И третий аспект – проведение кластерного анализа можно производить на постоянной основе, с использованием одной и той же выборки и системы показателей за разные временные периоды. Таким образом, мы получаем представление о динамике развития районов, траектории их движения относительно уже сформировавшихся групп, а это может служить основой для дальнейшего прогнозирования.

Исходя из этого, представим схему оценки социально-экономического положения муниципальных районов для определения направлений государственного регулирования (рис. 1).



**Рисунок 1. Этапы оценки социально-экономического положения муниципальных районов**  
**Figure 1. Stages of assessment of the socio-economic situation of municipal districts**

Применение кластерного анализа предполагает следующую последовательность [16]:

1. Обоснование выборки для кластеризации. В схеме оценки социально-экономического положения района выборкой для кластеризации будет вся совокупность районных показателей каждого отдельного муниципалитета.

2. Определение множества переменных, по которым будут производиться оценка объектов в выборке. Прежде, чем выделять конкретные переменные для оценки социально-экономического состояния муниципальных районов стоит отметить, что наиболее подходящим методом оценки будет нисходящая кластеризация, подразумевающая, что все объекты (регионы) принадлежат одному кластеру (Ивановской области). В ходе итеративного процесса будет проводиться разделение кластеров на подкластеры, образуя, таким образом, дендрограмму. Первоначальное и самое логичное разделение всех районов на 3 группы – развивающиеся, стагнирующие и депрессивные. Это дробление представляет собой не что иное, как выделение первого уровня подкластеров. Среди переменных, по которым будут оцениваться объекты в выборке, могут быть: количество трудоспособного населения, выручка крупных организаций на территории района, количество организаций, уровень инвестиций в основной капитал и т.д.

3. Вычисление значений меры сходства между разными объектами. Для социальных наук наибольшее распространение получило использование коэффициентов корреляции и меры расстояния при оценке сходства и различий.

4. Применение метода кластерного анализа для создания групп сходных объектов. Сам кластерный анализ достаточно трудоёмкий, поэтому для его проведения необходимо специальное программное обеспечение (в исследовании использована аналитическая платформа KNIME). Предполагается, что кластерный анализ позволит обосновать выделение новых кластеров муниципалитетов.

KNIME - это модульная платформа с открытым исходным кодом, предназначенная для анализа данных и составления отчетов. KNIME интегрирует различные компоненты для машинного обучения и интеллектуального анализа данных с помощью модульной концепции конвейера данных. Среда KNIME позволяет пользователю визуально создавать поток данных, выборочно выполнять анализ шагов, а затем исследовать результаты посредством интерактивного просмотра данных и моделей [15].

5. Проверка достоверности результатов кластерного решения. Для проверки достоверности предполагается использовать тесты значимости для признаков, используемых при создании кластеров, и повторные выборки.

При анализе данных по районам территориальных образований бывает очень сложно выделить основные характеристики. Классификация муниципальных образований обычно проводится по двум группам показателей:

1) Характеризующим социально-экономическое развитие. К ним относят: индекс физического объема валового муниципального продукта; индекс промышленного производства; индекс производства продукции сельского хозяйства; объем инвестиций в основной капитал; оборот розничной торговли; среднемесячную начисленную заработную плату работников; среднедушевые денежные доходы населения в месяц; просроченную задолженность по заработной плате; уровень зарегистрированной безработицы; объем работ по статье «Строительство»; строительство жилых домов;

2) Отражающим демографическое положение. Традиционно к ним относятся: площадь территории; численность и плотность населения; коэффициент естественного прироста (убыли) населения на 1000 чел.; миграционный прирост (убыль) на 1000 человек населения.

При наличии большого количества данных по каждому отдельному району и при большом количестве муниципальных образований в рамках региона выделить какие-то закономерности бывает затруднительно. В связи с этим предлагается использовать для анализа метод главных компонент (РСА – principal component analysis), который предполагает создание некоей упрощенной модели, которая позволяет отразить главное [19]. При использовании данного метода предполагается, что часть

признаков сильно зависят друг от друга, и нет необходимости использовать избыточные данные, когда можно выделить только основные.

Для того чтобы произвести оценку позиций районов Ивановской области с помощью интеллектуального анализа необходимо проведение поэтапного кластерного анализа на базе 6 статистических показателей. Их выбор обусловлен полнотой и достоверностью информации, характеризующей соответствующие условия развития регионов.

Районы Ивановской области существенно различаются по площади и количеству населения, проживающего в них. Так, самым маленьким по численности населения является Верхнеландеховский район (4339 чел.), а самым заселенным – городской округ Иваново (405053 чел.). Соотношение составляет примерно 1:90. Примерно такую же ситуацию можно наблюдать и по большинству других показателей, характеризующих уровень социально-экономического развития районов области. Вследствие этой неоднородности появляется возможность выделения среди всей совокупности муниципальных территорий определенных групп (кластеров).

Для анализа будем использовать файл с исходными данными по районам Ивановской области. Файл содержит несколько характеристик района: население (Н), трудоспособное население (ТН), инвестиции в основной капитал (ИОК), выручка крупных и средних организаций (ВО), количество организаций (КО), оборот розничной торговли (ОРТ). Общее количество муниципальных районов – двадцать семь.

Row ID	S Район	I Населе...	I Трудос...	I Инвест...	S Выруч...	S Количе...	D Обор...
Row0	гор.округ Иваново	405053	228177	14743592	105299674	18271	46,039,035
Row1	гор.округ Вичуга	33794	17107	198213	596709	310	1,938,638
Row2	гор.округ Кинешма	81986	43020	873134	6454897	956	4,838,885
Row3	гор.округ Тейково	32033	18090	270256	828848	356	2,309,686
Row4	гор.округ Шуя	57569	30329	1213400	12327621	959	4,066,861
Row5	гор.округ Кохма	30220	17051	386851	909389	495	1,700,014
Row6	Вичугский мун. р/он	17083	8934	17574	447181	117	436,074
Row7	Верхнеландеховс...	4339	2324	30044	379596.861...	56	110,760.703
Row8	Гаврилово-Поса...	15666	8305	667434	781538	173	437,865
Row9	Заволжский мун...	14553	7024	29045	81667	167	431,463
Row10	Ивановский мун. ...	37604	19988	1911188	16492157	1036	4,733,328
Row11	Ильинский мун. р...	8075	4037	31240	1082270	93	1,420,109.91
Row12	Кинешенский мун...	20916	11515	142548	5049966	243	381,949
Row13	Комсомольский м...	19723	10385	117428	138699	209	614,227
Row14	Лажанский мун. ...	18294	9659	335525	10187566	298	689,986
Row15	Луховский мун. р/он	7976	4078	72024	17581	92	628,981.11
Row16	Палехский мун. р...	9302	4590	39894	398745	120	314,890
Row17	Пестяковский му...	5649	2840	10240	26849	74	222,273
Row18	Приволжский мун...	23497	12561	139499	1378695	263	1,112,326
Row19	Пучежский мун. р...	10480	4635	269423	265711	128	405,162
Row20	Родниковский му...	32704	16805	494395	2566074	308	1,358,784
Row21	Савинский мун. р...	10580	5157	24896	39328	126	276,690
Row22	Тейковский мун. ...	10647	5131	34739	344928	131	895,614.406
Row23	Фурмановский му...	39560	20770	600345	2531447	300	1,812,415
Row24	Шуйский мун. р/он	21333	11232	9370	24819	254	1,736,534.802
Row25	Южский мун. р/он	22546	12921	106712	696122	173	710,218
Row26	Юрьевский мун....	12998	6405	58211	62285	123	425,472

Рисунок 2. Преобразованный файл с исходными данными в KNIME Analytics  
 Figure 2. Converted file with source data in KNIME Analytics

Рассмотрим показатели, которые позволяют рассчитать KNIME Analytics. Первый показатель – это матрица ковариации.

Ковариационная матрица представляет собой квадратную матрицу для понимания взаимосвязей, представленных между различными переменными в наборе данных. Исходя из этого, можно заметить, что, например, показатели населения и трудоспособного населения изменяются однонаправленно. Также рост инвестиций в основной капитал соответствует увеличению выручки организаций, осуществляющих деятельность на территории района. Таким образом, чем больше сумма инвестиций в район, тем больше объемы розничной торговли и выручка организа-

ций (рис. 3). Логика понятна, как и явная взаимосвязь количества организаций на территории района с трудоспособным населением.

Полученные данные возможно визуализировать с помощью инструментов Scatter Matrix (рис. 4) и Scatter Plot (рис. 5).

Согласно рис. 5 можно говорить о наличии явной зависимости между главными компонентами, что делает актуальным применение данного метода для анализа предложенных данных.

Для проведения кластеризации будем использовать несколько наиболее известных методов: иерархическая кластеризация, метод k-средних, DBSCAN [10, 19, 20]. На платформе KNIME процесс анализа будет выглядеть следующим образом (рис. 6).

Row ID	D Население,...	D Трудоспос...	D Инвестиции ...	D Выручка крупн...	D Количест...	D Оборот розни...
Население, ч...	5,695,476,629....	3,211,659,989....	209,127,656,850...	1,494,426,538,829....	258,368,064....	652,417,455,264....
Трудоспособ...	3,211,659,989....	1,811,606,299....	118,119,176,187...	843,984,523,985.44	146,008,234....	368,362,750,722....
Инвестиции ...	209,127,656,85...	118,119,176,18...	7,905,351,699,8...	56,537,203,100,238...	9,715,628,06...	24,382,047,756,6...
Выручка кру...	1,494,426,538,...	843,984,523,98...	56,537,203,100,...	409,220,580,993,10...	69,410,055,5...	174,293,166,713,...
Количество ...	258,368,064.416	146,008,234.101	9,715,628,061.682	69,410,055,505.063	12,046,453.678	30,124,708,366.092
Оборот розн...	652,417,455,26...	368,362,750,72...	24,382,047,756,...	174,293,166,713,45...	30,124,708,3...	75,840,923,142,2...

Рисунок 3. Матрица ковариации исходных данных  
Figure 3. Covariance matrix of source data

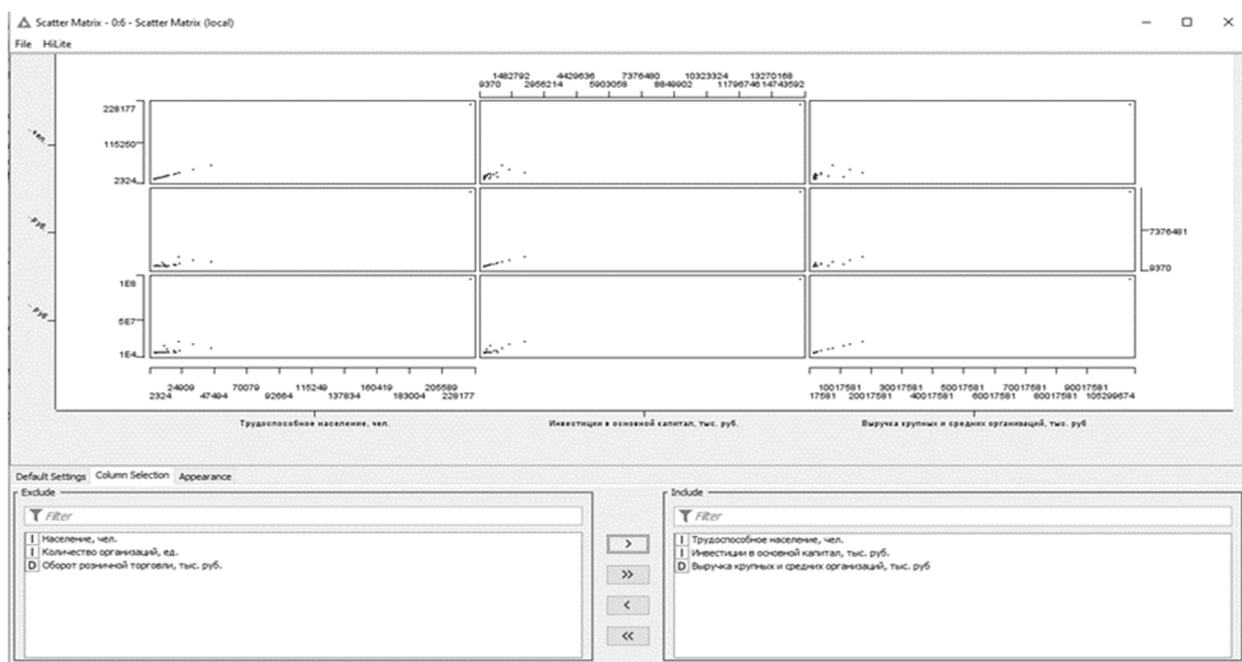


Рисунок 4. Матрица рассеяния  
Figure 4. Scatter matrix

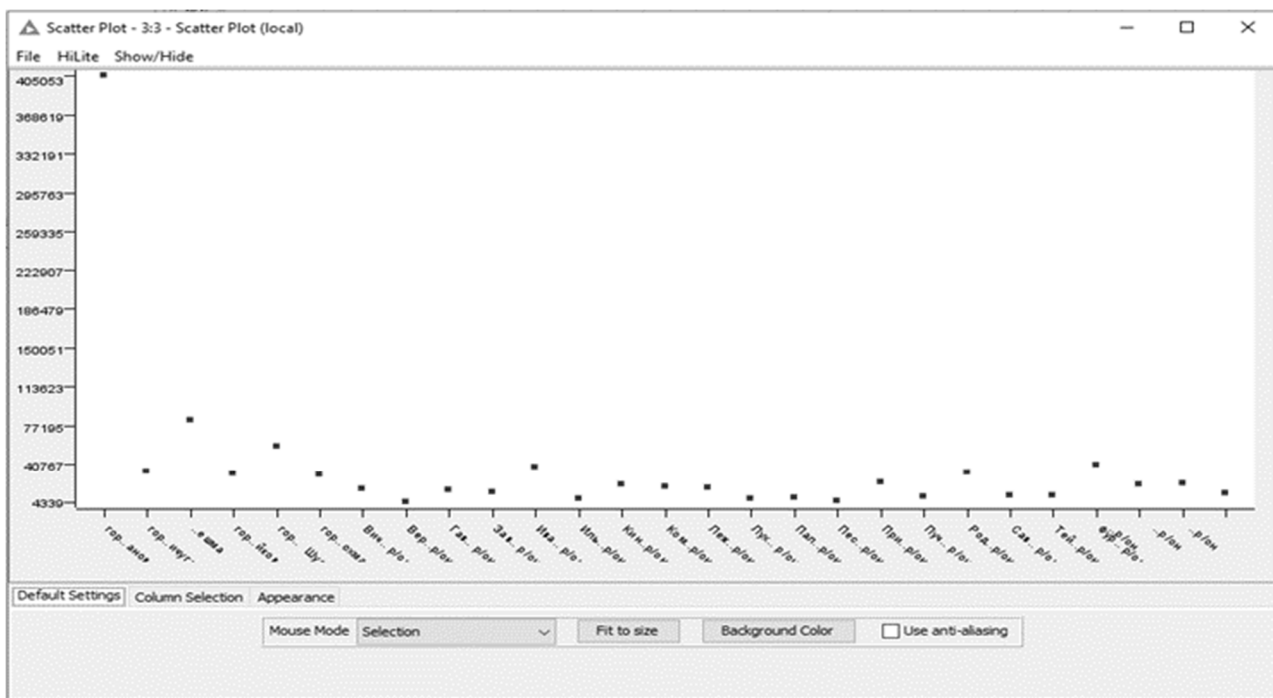


Рисунок 5. Точечная диаграмма  
Figure 5. Point diagram

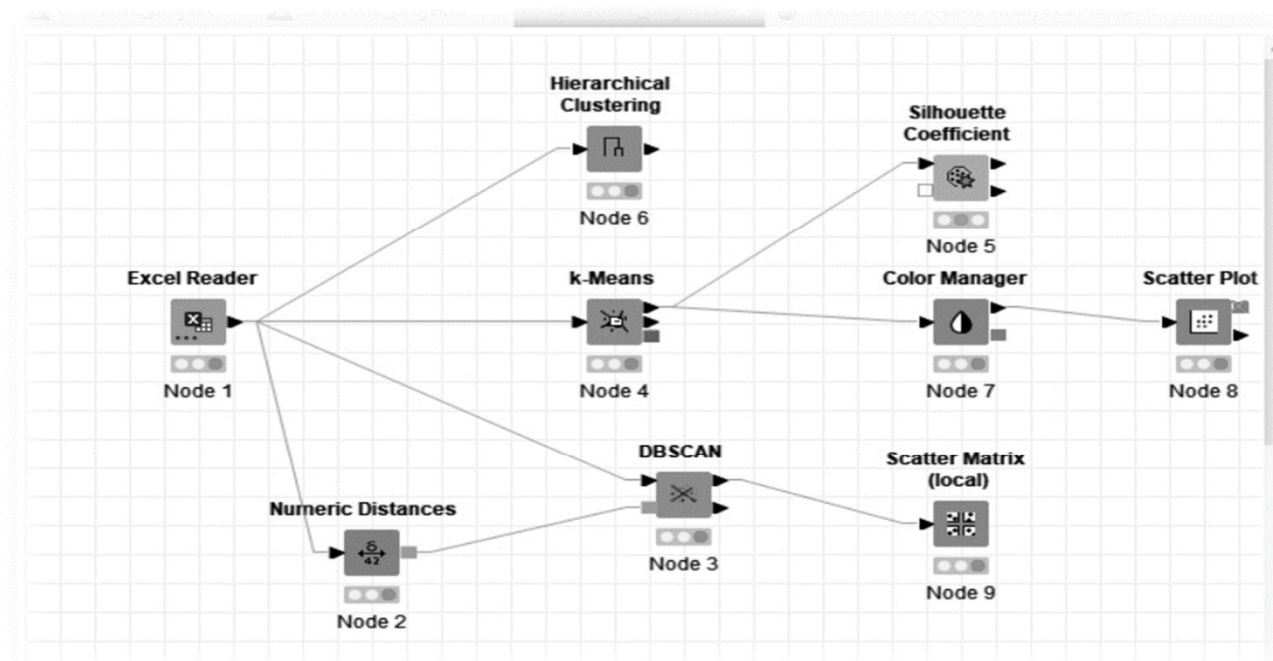
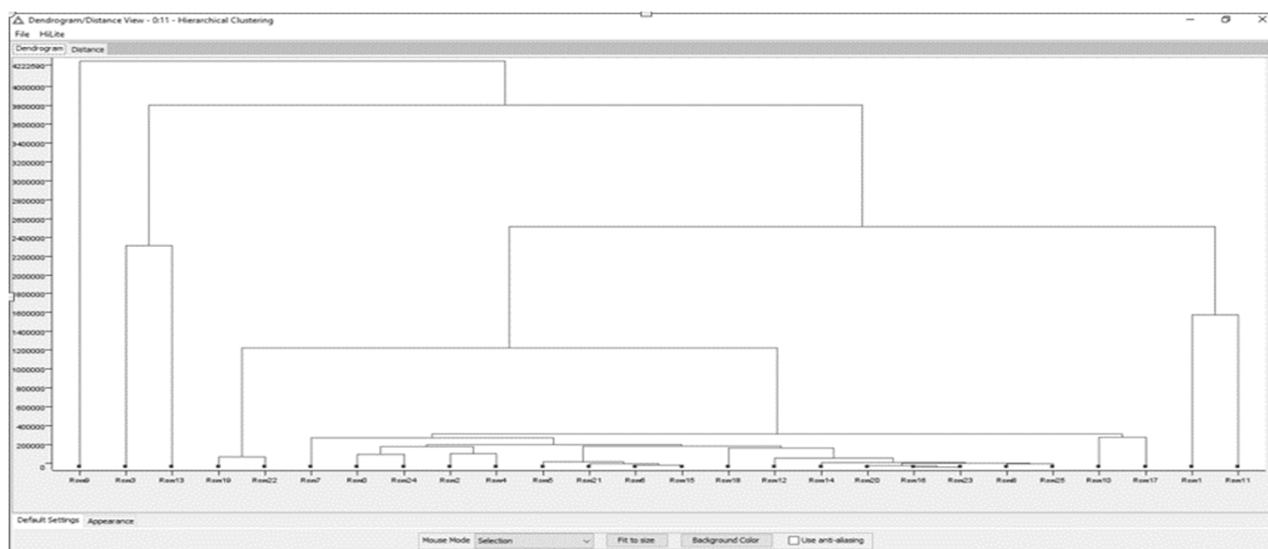


Рисунок 6. Процесс проведения анализа данных с помощью кластеризации  
Figure 6. The process of analyzing data using clustering

Иерархическая кластеризация – это тип неконтролируемого алгоритма машинного обучения, используемого для кластеризации немаркированных точек данных. Подобно K-means clustering, иерархическая кластеризация также группирует вместе точки данных с аналогичными

характеристиками. В некоторых случаях результат иерархической кластеризации и кластеризации K-средних может быть сходным [17].

С помощью дендрограммы можно предположить, какие объекты попали в какие кластеры (рис. 7).



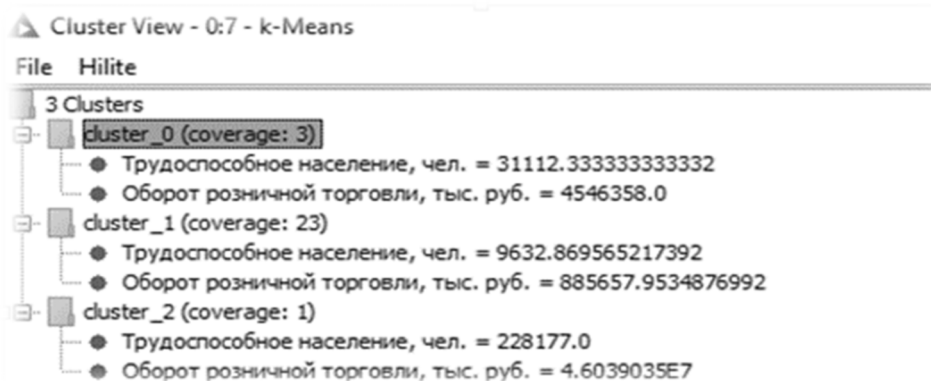
**Рисунок 7. Дендрограмма**  
**Figure 7. Dendrogram**

Основной вывод, который можно сделать согласно полученным данным, это предположить наличие 3 кластеров, объединяющих исходные данные со схожими характеристиками. Это предположение будем в дальнейшем использовать при применении в анализе алгоритма k-средних.

При анализе данных о районах (исходя из выше представленной дендрограммы) построим алгоритм k-means на основе предположения о 3 кластерах (сильные, средние и слабые по своему социально-экономическому развитию группы).

Данные удобнее анализировать в двухмерной плоскости, поэтому для анализа можно брать попарно показатели, которые представляют интерес.

Первой парой показателей будут трудоспособное население и оборот розничной торговли. При работе алгоритма k-means было выделено 3 кластера, по количеству – 3, 23 и 1 единиц в каждом кластере (рис. 8). Наличие отдельного кластера с единицей говорит о том, что один из районов обладает характеристиками весьма отличными от других. Использование данных по городскому округу Иваново носит спорный характер, и, возможно, его нужно было исключить из кластеризации. Попробуем выделить 4 кластера, предполагая, что один из них будет представлен городским округом Иваново.



**Рисунок 8. Кластеризация данных по параметрам**  
**«Трудоспособное население» и «Оборот розничной торговли»**

**Figure 8. Clustering of data on the parameters «Working-age population» and «Retail trade turnover»**

Исходя из усредненных данных, можно сказать, что самый высокий оборот розничной торговли наблюдается в группе из 23 единиц, которая характеризуется средним количеством трудоспособного населения. Это подтверждает сделанный выше вывод о том, что деление районов исключительно по территориальному признаку

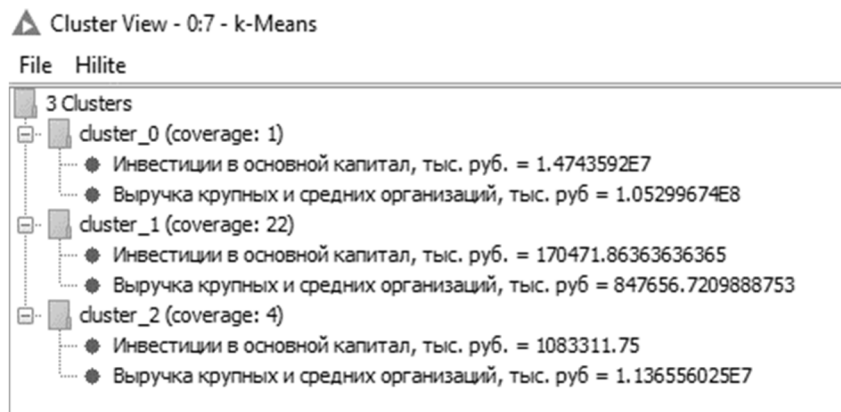
способствует значительному разрыву в социально-экономическом развитии районов.

Проведем кластеризацию данных по параметрам «Инвестиции в основной капитал» и «Выручка крупных и средних организаций» (рис. 9). Исходя из усредненных данных, можно сказать, что самая низкая выручка наблюдается в группе из 22 единиц, которая характеризуется низким

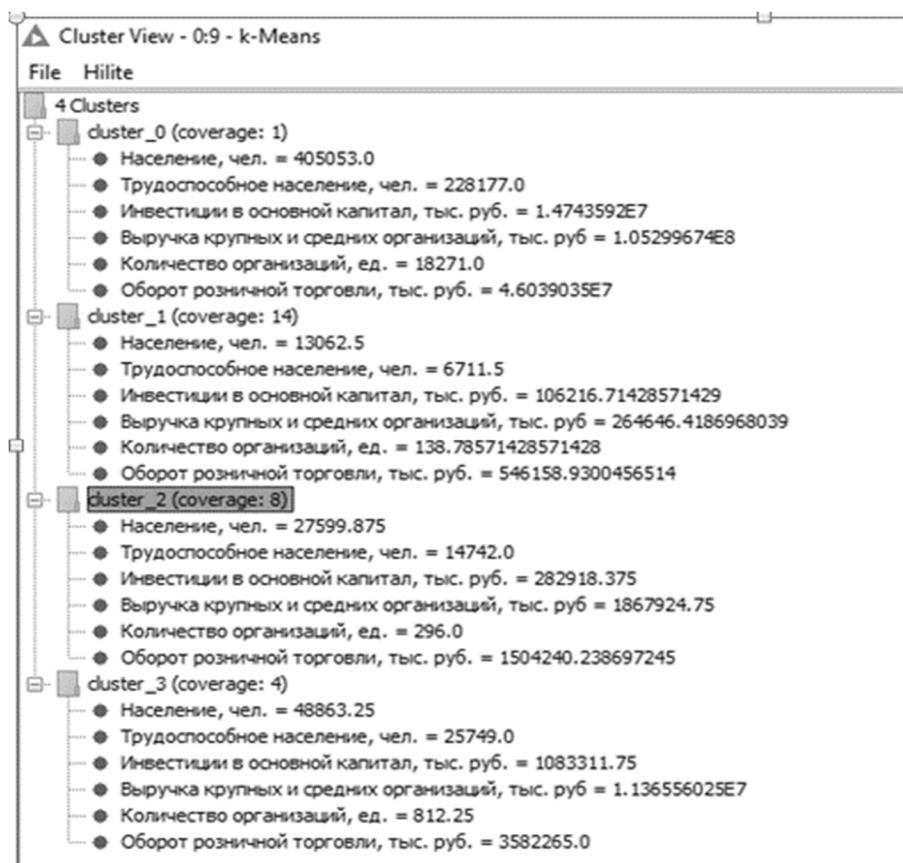
объемом инвестиций. В районах со средним объемом инвестиций - 4 единицы, при этом они генерируют выручку в среднем сопоставимую с районным лидером – городским округом Иваново. Это подтверждает выводы о том, что большинство

районов Ивановской области имеют слабое социально-экономическое положение.

Предположим, что количество кластеров равно 4, и будем использовать все показатели социально-экономического развития (рис. 10).



**Рисунок 9. Кластеризация данных по параметрам «Инвестиции в основной капитал» и «Выручка крупных и средних организаций»**  
**Figure 9. Clustering of data on the parameters «Investments in fixed assets» and «Revenue of large and medium-sized organizations»**



**Рисунок 10. Кластеризация данных на 4 кластера по всем параметрам социально-экономического развития районов**  
**Figure 10. Clustering of data into 4 clusters on all parameters of socio-economic development of districts**



▲ Labeled input - 0:9 - k-Means

File Edit Hilite Navigation View

Table "default" - Rows: 27 Spec - Columns: 8 Properties Flow Variables

Row ID	S Район	I Населе...	I Трудос...	I Инвест...	D Выруч...	I Количе...	D Оборо...	S Cluster
Row0	гор.округ Иваново	405053	228177	14743592	105,299,674	18271	46,039,035	cluster_0
Row1	гор.округ Вичуга	33794	17107	198213	596,709	310	1,938,638	cluster_2
Row2	гор.округ Кинешма	81986	43020	873134	6,454,897	956	4,838,885	cluster_3
Row3	гор.округ Тейково	32033	18090	270256	828,848	356	2,309,686	cluster_2
Row4	гор.округ Шуя	57569	30329	1213400	12,327,621	959	4,066,861	cluster_3
Row5	гор.округ Кохма	30220	17051	386851	909,389	495	1,700,014	cluster_2
Row6	Вичугской мун. р/он	17083	8934	17574	447,181	117	436,074	cluster_1
Row7	Верхнеландеховск...	4339	2324	30044	379,596.862	56	110,760.703	cluster_1
Row8	Гаврилово-Поса...	15666	8305	667434	781,538	173	437,865	cluster_1
Row9	Заволжской мун....	14553	7024	29045	81,667	167	431,463	cluster_1
Row10	Ивановской мун. ...	37604	19988	1911188	16,492,157	1036	4,733,328	cluster_3
Row11	Ильинской мун. р...	8075	4037	31240	1,082,270	93	1,420,109.91	cluster_2
Row12	Кинешемской мун...	20916	11515	142548	5,049,966	243	381,949	cluster_2
Row13	Комсомольской м...	19723	10385	117428	138,699	209	614,227	cluster_1
Row14	Лежневской мун. ...	18294	9659	335525	10,187,566	298	689,986	cluster_3
Row15	Луховской мун. р/он	7976	4078	72024	17,581	92	628,981.11	cluster_1
Row16	Палеховской мун. р...	9302	4590	39894	398,745	120	314,890	cluster_1
Row17	Пестяковской му...	5649	2840	10240	26,849	74	222,273	cluster_1
Row18	Приволжской мун...	23497	12561	139499	1,378,695	263	1,112,326	cluster_2
Row19	Пучежской мун. р...	10480	4635	269423	265,711	128	405,162	cluster_1
Row20	Родниковской му...	32704	16805	494395	2,566,074	308	1,358,784	cluster_2
Row21	Савинской мун. р...	10580	5157	24896	39,328	126	276,690	cluster_1
Row22	Тейковской мун. ...	10647	5131	34739	344,928	131	895,614.406	cluster_1
Row23	Фурмановской му...	39560	20770	600345	2,531,447	300	1,812,415	cluster_2
Row24	Шуйской мун. р/он	21333	11232	9370	24,819	254	1,736,534.802	cluster_1
Row25	Южновской мун. р/он	22546	12921	106712	696,122	173	710,218	cluster_1
Row26	Юрьевцевской мун....	12998	6405	58211	62,285	123	425,472	cluster_1

Рисунок 11. Распределение районов по 4 кластерам  
Figure 11. Distribution of districts into 4 clusters

Городской округ Иваново выделен в отдельный кластер, далее районы распределены следующим образом:

cluster\_1 – (14 районов) – низкие показатели социально-экономического развития;

cluster\_2 – (8 районов) – средние показатели социально-экономического развития;

cluster\_3 – (4 района) – высокие показатели социально-экономического развития.

Отдельно будем учитывать cluster\_0 (1 район – областной центр) – самые высокие показатели социально-экономического развития.

По всем анализируемым показателям городской округ (г.о.) Иваново является относительно благополучным с точки зрения социально-экономического развития муниципальным образованием. На его долю приходится более 41% жителей Ивановской области, что определяется относительно высокими показателями инфраструктурного и социально-экономического развития данной территории. Таким образом, в Ивановской области реализована центрально-периферийная модель пространственного развития.

В рамках проведенного исследования социально-экономического развития муниципальных образований были решены частные вопросы, связанные с изучением закономерностей социально-экономического развития муниципальных образований. Выделены кластеры поселений, в которых наблюдаются различные закономерности, влияющие на наблюдаемые параметры. Построены карты параметров, дающие возможность оценки ситуации и сравнения с другими муниципалитетами. На наш взгляд для более тщательного и точного анализа влияния параметров (а также управляющих воздействий) необходима более глубокая динамика данных (5-10 лет), более детальное фиксирование изменений параметров и включение в модели параметров «внешнего» окружения. Для увеличения точности моделей необходимо, в первую очередь, создавать и использовать системы мониторинга показателей социально-экономического развития муниципальных образований [9], а также решать задачу накопления информации об индикаторах в формате OLAP. Результаты исследований также имеют большое значение для фи-

нансовой сферы, поскольку позволят прогнозировать денежные и финансовые показатели в территориальном разрезе [3, 4, 5, 7].

#### ЗАКЛЮЧЕНИЕ

На основе полученных данных можно сделать следующие выводы:

1. Экономическая активность концентрируется вокруг региональной столицы. Она стимулирует развитие агломерации, поскольку экономические и социальные эффекты, создаваемые г.о. Иваново, начинают распространяться и на соседние муниципальные образования.

2. Кроме г.о. Иваново относительно благополучные показатели развития характерны для других муниципальных образований: г.о. Кинешма, г.о. Шуя, Ивановский и Лежневский муниципальные районы. Полученные результаты кластеризации позволяют сделать вывод о реализации региональными властями в Ивановской области политики стимулирующего развития. Развитие всех этих муниципальных образований состоялось благодаря реализации на их территориях крупных инвестиционных проектов. Стягивание экономической активности в отдельные муниципалитеты сдерживает развитие других территорий в силу ограниченности трудовых ресурсов и инвестиций. На фоне отрицательных значений показателя естественного прироста во всех муниципальных образованиях в долгосрочной перспективе проблема пространственных дисбалансов будет только усугубляться.

3. Периферийные территории испытывают дефицит социального капитала, в связи с

чем требуется баланс между стимулирующей и выравнивающей политикой регионального развития Ивановской области. Усиление выравнивающей составляющей региональной политики повышает риски для региональной бюджетной системы, потому необходимо наращивать собственные доходные источники.

4. Депрессивные муниципальные образования характеризуются стагнацией базовых отраслей (низкие производительность труда, инвестиции, налоговый потенциал), неудовлетворительной демографической структурой, слабыми инфраструктурами. В отношении этих территорий приоритетными должны стать меры сглаживающей региональной политики, связанные с ростом доходов, занятых в бюджетном секторе, а также реализация агропищевого, природоохранного и рекреационного потенциалов муниципальных образований.

5. Депопуляция территорий фрагментирует пространство Ивановской области. Поддерживать слабозаселенные территории можно за счет локальных проектов, повышающих динамику экономического роста, улучшающих экономико-географическое положение муниципальных образований за счет инфраструктурных проектов, на реализацию которых необходимы государственные инвестиции.

*Статья подготовлена по результатам исследований, выполненных в рамках гранта по договору с Ивановским государственным университетом N 05-21-Д от 16.04.2021 г.*

#### ЛИТЕРАТУРА

1. Национальная программа «Цифровая экономика Российской Федерации», утв. Президиумом Совета при Президенте РФ по стратегическому развитию и национальным проектам. Протокол от 04.06.2019 №7.
2. **Валинурова А.А.** Теоретическое обоснование использования метода кластеризации в банковском резервировании. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2021. №2(48). С. 17-22.
3. **Валинурова А.А., Данилова С.В.** Новый подход к формированию резервов в коммерческом банке. *Вестник Ивановского государственного университета. Серия: Экономика*. 2020. №4(46). С. 93-97.
4. **Валинурова А.А., Смирнова Е.М., Валинуров Т.Р., Бальясова Е.Д.** Трансформация подходов к оценке эффективности отделения банка в условиях цифровой экономики. *Вестник Ивановского государственного университета. Серия: Экономика*. 2020. №2(44). С. 90-97.
5. **Валинурова А.А., Смирнова Е.М., Ксенофонтова О.Л.** Интеллектуальное дистанционное банковское обслуживание и его особенности. *Современные наукоемкие технологии. Региональное приложение*. 2021. №2(66). С. 16-21.
6. **Данилова С.В.** Актуальные проблемы бизнес-аналитики и интеллектуального анализа данных в условиях цифровой экономики. В сб.: «Научно-исследовательская деятельность в классическом университете: традиции и инновации». Матер. Межд. н.-пр. фестиваля. 2020. С. 386-388.

#### REFERENCES

1. National Program «Digital Economy of the Russian Federation», approved by the Presidium of the Presidential Council for Strategic Development and National Projects, Protocol of 04.06.2019 N7. (in Russian).
2. **Valinurova A.A.** Theoretical substantiation of the use of clustering method in bank reservation. *Ivecofin*. 2021. N2(48). P. 17-22. (in Russian).
3. **Valinurova A.A., Danilova S.V.** A new approach to the formation of reserves in a commercial bank. *Bulletin of the Ivanovo State University. Series: Economics*. 2020. N4(46). P. 93-97. (in Russian).
4. **Valinurova A.A., Smirnova E.M., Valinurov T.R., Balyasova E.D.** Transformation of approaches to assessing the effectiveness of a bank branch in the digital economy. *Bulletin of the Ivanovo State University. Series: Economics*. 2020. N2(44). P. 90-97. (in Russian).
5. **Valinurova A.A., Smirnova E.M., Ksenofontova O.L.** Intelligent remote banking and its features. *Modern high-tech technologies. Regional application*. 2021. N2(66). P. 16-21. (in Russian).
6. **Danilova S.V.** Actual problems of business analytics and data mining in the digital economy. *Materials of the scientific and practical Festival «Research activity in the classical University: traditions and innovations»*. 2020. P. 386-388. (in Russian).

7. **Дубова С.Е.** Анализ потенциала распространения феномена цифровых денег центральных банков кредитными организациями и институтами нефинансового сектора экономики. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2021. №3(49). С. 59-68.
8. **Дугушкина Н.В.** Обзор популярных методов кластеризации в машинном обучении. *Наукофера*. 2020. №7. С. 112-118.
9. **Камаев В. А., Щербаков М. В., Аль-Катабери А.С.** Анализ социально-экономического развития муниципальных образований на примере Волгоградской области с использованием моделей data mining. *Изв. ВолгГТУ. Серия «Актуальные проблемы управления, вычислительной техники и информатики в технических системах»*. 2010. Т. 6. №8. С. 103-106.
10. **Кревский М.И., Сметанев Д.М.** Применение машинного обучения и process mining к задачам кластеризации процессов в финансовой сфере. *Современные инструменты, методы и технологии управления знаниями*. 2020. №3. С. 52-60.
11. **Кузнецов Д.Ю., Трошина Т.Л.** Кластерный анализ и его применение. *Ярославский педагогический вестник*. 2006. №4(49). С. 103-107. <https://cyberleninka.ru/article/n/klasternyy-analiz-i-ego-primenenie/viewer>.
12. **Мерзликина Г.С., Кузьмина Е.В.** Методика оценки эффективности деятельности кластера на основе сбалансированной системы показателей в условиях цифровой экономики. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2020. №2(44). С. 14-23.
13. Минцифры России будет координировать развитие новых производственных и коммуникационных технологий. <https://digital.gov.ru/ru/events/41196/>
14. **Миркин Б.Г.** Методы кластер-анализа для поддержки принятия решений. М.: НИУ ВШЭ. 2011. 88 с.
15. Официальный сайт платформы KNIME. <https://www.knime.org>.
16. **Тюрин А.Г., Зуев И.О.** Кластерный анализ, методы и алгоритмы кластеризации. *Вестник МГТУ МИРЭА*. 2014. №2(3). С. 86-97.
17. **Усман М.** Иерархическая кластеризация с помощью Python и Scikit-Learn. <https://pythobyte.com/hierarchical-clustering-with-python-and-scikit-learn-768953fe>.
18. **Хомякова А.А., Мизгирев Л.С., Шергин В.В.** Использование методов интеллектуального анализа данных в процессах управления инвестиционной привлекательностью региона. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2020. №2(44). С. 14-23.
19. **Gorban A., Kegl B., Wunsch D., Zinovyev A.** Principal Manifolds for Data Visualization and Dimension Reduction. Berlin – Heidelberg – New York. 2007.
20. **Kanungo T.** An efficient k-means clustering algorithm: Analysis and implementation. *Pattern Analysis and Machine Intelligence*. 2002. Vol. 24. N 7. P. 881–892.
7. **Dubova S.E.** Analysis of the potential for the spread of the phenomenon of digital money of central banks by credit organizations and institutions of the non-financial sector of the economy. *Ivecofin*. 2021. N3(49). P. 59-68. (in Russian).
8. **Dugushkina N.V.** Review of popular clustering methods in machine learning. *Scienceasphere*. 2020. N7. P. 112-118. (in Russian).
9. **Kamaev V. A., Shcherbakov M. V., Al-Kataberi A.S.** Analysis of socio-economic development of municipalities on the example of the Volgograd region using data mining models. *Bulletin VSTU. Series «Actual problems of management, computer technology and informatics in technical systems»*. 2010. Vol. 6. N8. P. 103-106. (in Russian).
10. **Krevsky M.I., Smetanov D.M.** Application of machine learning and process mining to the tasks of clustering processes in the financial sector. *Modern tools, methods and technologies of knowledge management*. 2020. N3. P. 52-60. (in Russian).
11. **Kuznetsov D.Yu., Troshina T.L.** Cluster analysis and its application. *Yaroslavl Pedagogical Bulletin*. 2006. N4(49). P. 103-107. <https://cyberleninka.ru/article/n/klasternyy-analiz-i-ego-primenenie/viewer>. (in Russian).
12. **Merzlikina G.S., Kuzmina E.V.** Methodology for assessing the effectiveness of cluster activities based on a balanced scorecard in the digital economy. *Ivecofin*. 2020. N2(44). P. 14-23. (in Russian).
13. The Ministry of Digital Development, Communications and Mass Communications of Russia will coordinate the development of new production and communication technologies. <https://digital.gov.ru/ru/events/41196> (in Russian).
14. **Mirkin B.G.** Cluster analysis methods for decision support. Moscow: HSE. 2011. 88 p. (in Russian).
15. The official website of the KNIME platform. <https://www.knime.org>. (in Russian).
16. **Tyurin A.G., Zuev I.O.** Cluster analysis, clustering methods and algorithms. *Bulletin of MSTU MIREA*. 2014. N2(3). P. 86-97. (in Russian).
17. **Usman M.** Hierarchical clustering using Python and Scikit-learn. <https://pythobyte.com/hierarchical-clustering-with-python-and-scikit-learn-768953fe>. (in Russian).
18. **Khomyakova A.A., Mizgirev L.S., Shergin V.V.** The use of data mining methods in the processes of managing the investment attractiveness of the region. *Ivecofin*. 2020. N2(44). P. 14-23. (in Russian).
21. **Gorban A., Kegl B., Wunsch D., Zinovyev A.** Principal Manifolds for Data Visualization and Dimension Reduction. Berlin – Heidelberg – New York. 2007.
22. **Kanungo T.** An efficient k-means clustering algorithm: Analysis and implementation. *Pattern Analysis and Machine Intelligence*. 2002. Vol. 24. N 7. P. 881–892.

Поступила в редакцию 01.10.2021  
Принята к опубликованию 15.10.2021

Received 01.10.2021  
Accepted 15.10.2021