

УДК 519.25, 331.52

DOI: 10.22213/2410-9304-2025-2-80-92

## Информационные технологии обработки данных для анализа показателей рынка труда в разрезе по субъектам Российской Федерации

Д. Д. Вавилова, кандидат технических наук, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия

К. В. Кетова, доктор физико-математических наук, профессор,

ИжГТУ имени М. Т. Калашникова, Ижевск, Россия

К. М. Давлетшина, студент, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия

*Статья посвящена изучению и систематизации информационных технологий обработки статистических данных в применении к анализу рынка труда. Рассмотрены методы, которые способствуют пониманию структуры занятости и других аспектов рынка труда в РФ. Акцент сделан на систематизации методов обработки данных и их использовании для описательной статистики, кластеризации и геоаналитики показателей рынка труда в разрезе субъектов РФ. Результатом исследования является анализ характеристик рынка труда субъектов, статистический и корреляционный анализ данных, кластеризация субъектов РФ по показателям рынка труда, а также создание интерактивной карты, построенной по уровню занятости. Выполненная кластеризация определила группы субъектов со схожими показателями рынка труда по уровню занятости и безработицы, уровню рабочей силы, отношению среднедушевых денежных доходов населения к их средней величине по стране, отношению среднемесячной заработной платы к прожиточному минимуму, отношению удельного валового регионального продукта в субъекте к его средней величине по стране, а также по структуре занятого населения в основных видах экономической деятельности. Получено, что для 2023 года на рынке труда РФ формируются 4 кластера с субъектами, показатели которых внутри каждого из кластеров имеют близкие по величине значения. В первый кластер входит 22 субъекта РФ с низким уровнем занятости (55,4 %) и высоким уровнем безработицы (7,2 %) населения. Второй кластер включает в себя 3 субъекта и характеризуется высоким показателем занятости населения (73,1 %), высокой долей участия в рабочей силе (75,4 %), максимальным среди кластеров отношением среднедушевых денежных доходов населения к их средней величине по РФ (281,8 %). Почти четверть населения этих субъектов занята в добывающей отрасли. В третий кластер входит 9 субъектов РФ с более высокими показателями занятости населения (66,1 %) и долей участия в рабочей силе (68,0 %) по сравнению со средними значениями в РФ, а также минимальным среди кластеров уровнем безработицы. Четвертый кластер является наиболее обширным: он включает остальные 53 субъекта федерации. Его можно охарактеризовать низкими значениями рассматриваемых показателей. В этих субъектах уровень безработицы составляет всего 2,8 %.*

**Ключевые слова:** информационные технологии, анализ данных, кластеризация, рынок труда, субъект.

### Введение

Цифровизация различных сфер деятельности, которая является ключевым элементом построения цифровой экономики, основывается на информационных технологиях обработки данных [1]. Они позволяют работать с большими объемами информации, что актуально для нашей страны, поскольку территориальное деление увеличивает объемы статистической информации.

Технологии обработки данных интересуют исследователей. Так, в публикации [2] предложена модель сквозного процесса обработки и контроля данных при управлении организацией. В статье рассмотрены варианты совместного параллельного использования алгоритмов машинного обучения и искусственного интеллекта.

В исследовании [3] уделяется внимание использованию современных технологий обработки данных для решения проблемы создания ком-

фортной городской среды. Это методы цифрового анализа, блокчейн-технологий, а также интернета вещей.

В научной статье [4] описаны возможности применения интеллектуального анализа данных для решения задач в финансовой, маркетинговой, торговой и страховой деятельности компаний.

Применение основных методов анализа данных – регрессия, кластеризация, классификация – предоставляет возможность выявления зависимостей в наборе изучаемых показателей.

Также информационные технологии помогают оценивать состояние, анализировать динамику и строить прогноз развития сложных социально-экономических систем. К таким системам можно отнести рынок труда. В работе [5] рынок труда рассматривается в виде динамической системы потоков и накопителей; авторы указали основные подходы к моделированию состояния и движения кадров на рынке труда РФ.

Другие исследователи в работе [6] изучают приемы имитационного моделирования в целях принятия решений на основе анализа состояния социально-экономических систем. В исследовании [7] представлен зарубежный опыт применения технологий анализа больших данных для повышения точности и качества аналитики рынка труда.

Таким образом, применение информационных технологий для анализа рынка труда является актуальной задачей. Оперативный анализ показателей рынка труда (динамики трудовых ресурсов, уровня занятости, безработицы, кадровых потребностей и др.) выступает одной из наиболее сложных и важных задач современной цифровой экономики в РФ. Мониторинг состояния и тенденций изменения показателей рынка труда в субъектах федерации необходим для построения надежного прогноза развития ситуации в РФ в целом.

Целью настоящего исследования является системное изучение современных инструментов информационных технологий обработки данных и их применение для анализа показателей рынка труда в разрезе по субъектам РФ.

В настоящее время в РФ 89 субъектов. Если мы рассмотрим предоставленный Росстатом список регионов (субъектов, областей) РФ в 2025 году<sup>1</sup>, то там присутствует 91 позиция. Выделяется Тюменская область с Ханты-Мансийским и Ямало-Ненецким автономными округами и Архангельская область с Ненецким автономным округом.

В настоящей работе рассмотрены 87 субъектов, поскольку информация по ряду регионов на данный момент пока отсутствует (Донецкая и Луганская республики, Херсонская и Запорожская области).

В работе используем программу Jupyter Notebook на языке программирования Python. Результатом применения информационных технологий обработки данных является отражение в статье описательной статистики для исследуемых показателей рынка труда РФ, распределение субъектов федерации по кластерам в зависимости от выполненного анализа, а также создание интерактивной карты с распределением уровня занятости по субъектам РФ.

Описательная статистика позволяет охарактеризовать исследуемый набор данных для выявления взаимосвязей.

Распределение субъектов РФ по кластерам на основании показателей рынка труда в них необходимо для выявления общих тенденций и спе-

цифических характеристик, присущих группам субъектов со схожими проблемами и возможностями. Использование результатов кластеризации субъектов по показателям рынка труда помогает при разработке государственной политики занятости.

Можно внедрять дифференцированные подходы, ориентированные на конкретные потребности и потенциал субъектов каждого кластера, вместо универсальных мер, не учитывающих региональную специфику.

## Материалы

### и методы исследования

Методы обработки данных, применяемые в целях анализа показателей рынка труда, можно разделить на методы количественного и качественного анализа [8]. Важно сочетать несколько методов для получения более точных и полных выводов о состоянии и тенденциях изменения показателей рынка труда.

Количественный анализ позволяет выполнять мониторинг рынка труда, т. е. выявлять текущие тенденции, например, спроса на профессии, числа вакансий, определять наиболее востребованные специальности, анализировать уровень заработных плат, исследовать распределения размеров оплаты труда по отраслям, в том числе в разрезе по субъектам.

На основе выявленных тенденций появляется возможность предсказывать будущие изменения на рынке труда, к примеру, прогнозировать уровень безработицы и занятости, а также определять влияние различных процессов на них, например уровня цифровизации [9, 10].

Качественный анализ показателей рынка труда помогает прогнозировать и планировать, в каких отраслях экономики и в каком количестве нужны специалисты различных профессий и уровней квалификации.

Также качественный анализ показателей рынка труда способствует совершенствованию процесса планирования системы образования: для корректировки образовательных программ следует определить, какие навыки и профессии будут востребованы в будущем [11].

При анализе данных выделяют пять последовательных этапов [12, 13], такие как сбор, хранение, обработка, анализ визуализация.

На каждом этапе могут применяться различные информационные технологии. Рассмотрим основные современные инструменты информационных технологий обработки данных для анализа показателей рынка труда (табл. 1).

<sup>1</sup>Список регионов (субъектов, областей) РФ в 2025 году. URL: <https://www.statdata.ru/spisok-regionov-rossii-s-kodamy> (дата обращения: 02.02.2025).

**Таблица 1. Основные современные инструменты информационных технологий обработки данных для анализа показателей рынка труда**

**Table 1. Main modern tools of information technology data processing for the analysis of labour market in dicators**

№ п/п	Наименование этапа	Описание	Инструменты информационных технологий
1	Сбор данных	1.1. Источники данных: данные могут быть получены из государственных статистических служб, социальных сетей, систем учета, платформ для поиска работы, опросов работодателей	Госструктура Росстат; Платформа HeadHunter; Платформа Import.io; Инструмент ScrapingBot; Инструмент Octoparse; Инструмент DataOx; Инструмент Zyte; Программа Mozenda; API-инструмент ScraperAPI; API-инструмент Wintr; API-инструмент TestMace; API-инструмент RapidAPI; Прочие инструменты
		1.2. Веб-скрейпинг: автоматизированный процесс извлечения данных с веб-сайтов, например, вакансий, резюме, требований работодателей	
		1.3. Технология API (Application Programming Interface): использование API приложений для получения структурированной информации, например, данных о вакансиях, зарплатах и других параметрах	
2	Хранение данных	2.1. Базы данных: использование реляционных и нереляционных баз данных для хранения больших объемов информации	Система PostgreSQL; MySQL; Инструмент DBeaver; Программа MongoDB; Конструктор Firebird; Веб-сервис phpMyAdmin; Платформа Teamly; Yonote; Сервис Yandex Wiki
		2.2. Облачные хранилища: использование облачных платформ для хранения и обработки данных	
3	Обработка данных	3.1. Очистка данных: удаление дубликатов, несоответствий, ошибок и выбросов	Языки программирования Python, R; Язык запросов SQL; Программный продукт Microsoft Excel; Сервис YandexDataTransfer; Платформа SberDataPlatform; Технология ApacheHadoop; Платформа ApacheSpark; Модульная система Триафлай; Прочие инструменты
		3.2. Нормализация и стандартизация: приведение данных к единому масштабу	
		3.3. ETL-процессы: извлечение, трансформирование и загрузка данных (Extract, Transform, Load). Используется для приведения данных к единой структуре и загрузки их в хранилище данных	
4	Анализ данных	4.1. Описательная статистика: анализ средних значений, медиан, распределений и других показателей	Языки программирования Python, R; Язык запросов SQL; Программный продукт Microsoft Excel; Программа ClearAnalytics; Microsoft Power BI; Платформа Apache Spark; Платформа AnalyticWorkspace; Модульная система Триафлай; Облачная система VisaryBI; Сервис YandexDataLens; Прочие инструменты
		4.2. Прогнозирование: применение средств машинного обучения для прогнозирования исследуемых показателей	
		4.3. Кластеризация/ классификация: группировка данных для выявления схожих показателей	
		4.4. Текстовый анализ: анализ данных с использованием NLP (Natural Language Processing) для отражения содержания, анализа тональности	
5	Визуализация данных	5.1. Диаграммы, интерактивные дашборды: позволяют быстро оценить соотношение различных величин	Языки программирования Python, R; Microsoft Excel; Microsoft PowerBI; Программное обеспечение Tableau; BI-платформа YandexDataLens; Платформа Visiology; Платформа Datawrapper; Облачная система VisaryBI; Модульная система Триафлай; Прочие инструменты
		5.2. Геоаналитика: визуализация данных на картах, цветограммах	

Разграничение используемых инструментов при анализе показателей рынка труда в табл. 1 несколько условно, поскольку на разных этапах возможно применение одних и тех же информационных технологий. В частности, языки программирования (Python, R, SQL и др.) используются на этапах хранения, обработки, анализа и визуализации данных. В случае применения языка программирования Python для обработки и последующего анализа данных потребуется использование таких встроенных библиотек, как Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch.

Опишем подробнее этапы табл. 1 в рамках настоящего исследования, которое ограничивается представлением описательной статистики

Таблица 2. Используемые для анализа показатели рынка труда

Table 2. Labour market indicators used for analysis

№ п/п	Обозначение показателя	Наименование показателя	Пояснение к расчету показателя
1	X1	Уровень занятости населения, %	Отношение численности занятого в экономике населения к общей численности населения в возрасте 15 лет и старше, выраженное в процентах
2	X2	Уровень безработицы, %	Отношение численности безработного населения к общей численности населения в возрасте 15 лет и старше, выраженное в процентах
3	X3	Доля участия в рабочей силе, %	Отношение численности рабочей силы к общей численности населения в возрасте 15 лет и старше, выраженное в процентах
4	X4	Отношение среднедушевых денежных доходов населения к их средней величине по РФ, %	Отношение среднедушевых денежных доходов населения к их средней величине по стране за рассматриваемый год (средняя величина денежных доходов на одного жителя РФ в 2023 г. составила 49 805 руб./мес. <sup>8</sup> ), выраженное в процентах
5	X5	Отношение средней заработной платы к величине прожиточного минимума, %	Отношение средней номинальной начисленной заработной платы работников в целом в субъекте к величине прожиточного минимума, установленного для трудоспособного населения в целом по РФ за рассматриваемый год (величина прожиточного минимума в 2023 г. составляла 15 669 руб. <sup>9</sup> ), выраженное в процентах

<sup>2</sup> Рынок труда, занятость и заработная плата. Федеральная служба государственной статистики РФ. URL: [https://rosstat.gov.ru/labor\\_market\\_employment\\_salaries](https://rosstat.gov.ru/labor_market_employment_salaries) (дата обращения: 10.02.2025).

<sup>3</sup> Трудовые ресурсы, занятость и безработица. Федеральная служба государственной статистики РФ. URL: [https://rosstat.gov.ru/labour\\_force](https://rosstat.gov.ru/labour_force) (дата обращения: 10.02.2025).

<sup>4</sup> Статистика занятости и безработицы. Федеральная служба государственной статистики РФ. URL: <https://11.rosstat.gov.ru/storage/mediabank/Статистика%20занятости%20и%20безработицы.pdf> (дата обращения: 10.02.2025).

<sup>5</sup> Демография. Численность и состав населения. Федеральная служба государственной статистики РФ. URL: <https://rosstat.gov.ru/folder/12781> (дата обращения: 10.02.2025).

<sup>6</sup> Национальные счета. Валовой региональный продукт. Федеральная служба государственной статистики РФ. URL: <https://rosstat.gov.ru/statistics/accounts> (дата обращения: 10.02.2025).

<sup>7</sup> Регионы России. Социально-экономические показатели. Федеральная служба государственной статистики РФ. URL: <https://rosstat.gov.ru/folder/210/document/13204> (дата обращения: 10.02.2025).

<sup>8</sup> Национальные счета. Валовой региональный продукт. Федеральная служба государственной статистики РФ. URL: [https://rosstat.gov.ru/storage/mediabank/VRP\\_s1998.xlsx](https://rosstat.gov.ru/storage/mediabank/VRP_s1998.xlsx) (дата обращения: 10.02.2025).

<sup>9</sup> Величина прожиточного минимума. Федеральная служба государственной статистики РФ. URL: <https://rosstat.gov.ru/vpm> (дата обращения: 10.02.2025).

показателей рынка труда, кластеризацией субъектов РФ и визуализацией данных на картах.

#### Этап 1. Сбор статистических данных

Статистическая база настоящего исследования – это разделы «Рынок труда, занятость и заработная плата»<sup>2</sup>, «Трудовые ресурсы, занятость, безработица»<sup>3</sup>, «Статистика занятости и безработицы»<sup>4</sup>, «Демография»<sup>5</sup>, «Национальные счета»<sup>6</sup>, «Отраслевая структура занятости»<sup>7</sup> сайта Федеральной службы государственной статистики.

В работе в целях представления методологии анализа данных рассматривается 12 основных показателей рынка труда, представленных в табл. 2.

Окончание табл. 2

№ п/п	Обозначение показателя	Наименование показателя	Пояснение к расчету показателя
6	X6	Отношение валового регионального продукта в расчете на одного жителя в субъекте федерации к его средней величине по стране, %	Отношение валового регионального продукта в расчете на одного жителя в субъекте федерации к его средней величине по стране за рассматриваемый год (ВРП на одного жителя страны в 2023 г. составило 1 073,6 тыс. руб./год <sup>10</sup> )
7	X7	Удельный вес занятых в сельском, лесном хозяйстве, охоте, рыболовстве и рыбоводстве, %	Отношение численности занятых в сельском, лесном хозяйстве, охоте, рыболовстве и рыбоводстве к общей численности занятых в экономике данного региона за рассматриваемый год, выраженное в процентах
8	X8	Удельный вес занятых в добыче полезных ископаемых, %	Отношение численности занятых в добыче полезных ископаемых к общей численности занятых в экономике данного субъекта за рассматриваемый год, выраженное в процентах
9	X9	Удельный вес занятых в обрабатывающем производстве, %	Отношение численности занятых в обрабатывающем производстве к общей численности занятых в экономике данного субъекта за рассматриваемый год, выраженное в процентах
10	X10	Удельный вес занятых в строительстве, %	Отношение численности занятых в строительстве к общей численности занятых в экономике данного субъекта за рассматриваемый год, выраженное в процентах
11	X11	Удельный вес занятых в торговле оптовой и розничной, ремонте автотранспортных средств и мотоциклов, %	Отношение численности занятых в торговле оптовой и розничной, ремонте автотранспортных средств и мотоциклов к общей численности занятых в экономике данного субъекта за рассматриваемый год, выраженное в процентах
12	X12	Удельный вес занятых в области информации и связи, %	Отношение численности занятых в области информации и связи к общей численности занятых в экономике данного субъекта за рассматриваемый год, выраженное в процентах

### Этап 2. Хранение данных

База данных статистической информации, предназначенная для решения задачи анализа показателей рынка труда в разрезе по субъектам РФ, хранится в виде файла, реализованного в программном продукте Microsoft Excel.

### Этап 3. Обработка статистических данных

Обработка статистических данных выполняется с использованием программной среды Jupyter Notebook на языке программирования Python.

### Этап 4. Анализ данных

Как уже было сказано, в рамках настоящего исследования ограничимся представлением описательной статистики показателей рынка труда, кластеризацией субъектов РФ и визуализацией данных.

Описательная статистика включает в себя анализ ряда показателей математической статистики. Мы будем рассчитывать среднее значение, медиану, модальное значение, вариацию, дисперсию, среднеквадратическое отклонение, коэффициенты асимметрии и эксцесса [14].

Кластеризация предусматривает разбиение множества исследуемых субъектов РФ на однородные по показателям рынка труда группы (кластеры). Существуют различные алгоритмы кластеризации, например метод  $k$ -средних, иерархическая кластеризация, метод DBSCAN (Density-Based Spatial Clustering of Applications with Noise), метод Mean Shift, др. [15]. Каждый из них имеет свои особенности.

В настоящей работе для анализа показателей рынка труда в разрезе по субъектам РФ будем использовать метод  $k$ -средних. В этом методе вначале определяются центры кластеров, а далее к ним присоединяются последующие элементы на основе подсчета расстояния между объектами (используется принцип наименьшего расстояния). Итерации основаны на уменьшении общего квадратичного отклонения объектов кластеров от их центров.

Этот метод кластеризации требует заранее заданного числа кластеров  $k$ . Применим для определения оптимального значения числа кластеров «метод локтя» [16], суть которого заклю-

<sup>10</sup> Регионы России. Социально-экономические показатели. Федеральная служба государственной статистики РФ. URL: <https://rosstat.gov.ru/folder/210/document/13204> (дата обращения: 10.02.2025).

чается в рассмотрении объясненной дисперсии. В качестве наилучшего значения  $k$  принимается число, при котором прирост объясненной дисперсии  $S$  становится незначительным.

Представление описательной статистики и кластерный анализ реализуются на языке Python с помощью среды Jupyter Notebook и встроенных в нее библиотек: выполнения математических операций (Math); статистических операций (Statistics); научных вычислений (NumPy, Pandas); кластеризации данных (Scikit-learn).

#### Этап 5. Визуализация данных

Для визуализации показателей рынка труда в разрезе по субъектам РФ будем использо-

вать геоаналитическую систему отображения. Она представлена в виде интерактивной карты с отображением на ней показателей рынка труда.

На первоначальном этапе загружается файл с географическими данными и информацией о показателях рынка труда для каждого субъекта РФ. Географические данные представлены в формате GeoJSON.

Загружаем файл с данными и выводим часть информации (см. рис. 1).

	region	federal_district	population	geometry	employment
0	Пензенская область	Приволжский	1246609	MULTIPOLYGON (((42.08174 53.748, 42.08525 53.7...	65.1
1	Саратовская область	Приволжский	2404944	MULTIPOLYGON (((42.50367 51.66712, 42.50393 51...	65.1
2	Республика Мордовия	Приволжский	771373	MULTIPOLYGON (((42.16337 54.25811, 42.16397 54...	67.1
3	Тамбовская область	Центральный	966250	MULTIPOLYGON (((39.91569 52.70885, 39.92159 52...	63.0

Рис. 1. Частичное представление набора с географическими данными

Fig. 1. Partial representation of a geographic dataset

Каждая строка содержит информацию об одном субъекте РФ с указанием названия субъекта, федерального округа, численности населения, границы субъектов, а также уровня занятости населения.

Далее необходимо добавить на карту цветное отображение в зависимости от показателя уровня занятости населения. Нормализуем значения уровня занятости для цветовой шкалы и задаем цветовую палитру.

В цикле перебираем все субъекты из набора данных и создаем текстовое содержимое для каждого субъекта. Определяем цветовой индекс на основе нормализованного значения занятости в субъекте и отображаем полученную карту.

Построение интерактивной карты реализуется на языке программирования Python с помощью программной среды Jupyter Notebook и встроенных в нее библиотек: визуализации (Plotly, Matplotlib); обработки и анализа геоданных (GeoPandas); работы с геометрическими объектами (Shapely); с интерактивными картами (Ipyleaflet); создания интерактивных виджетов (Ipywidgets).

#### Результаты исследования:

##### описательная статистика,

##### кластеризация, геоаналитика

В ходе исследования были рассчитаны характеристики описательной статистики по показателям рынка труда (табл. 2) в разрезе субъектов РФ. Результаты представлены в табл. 3.

Таблица 3. Описательная статистика показателей рынка труда по субъектам РФ за 2023 г в %

Table 3. Descriptive statistics of labour market indicators by subjects of the Russian Federation for 2023, %

Характеристика	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Минимум	49,1	1,5	51,1	52,2	228,6	16,9	0,2	0,0	1,0	5,2	5,4	1,0
Максимум	77,8	27,8	79,1	312,8	1015,2	1117,3	20,7	24,9	24,0	16,8	23,4	5,7
Размах вариации	28,7	26,3	28,0	260,6	786,6	1100,3	20,5	24,9	23,0	11,6	18,0	4,7
Средняя арифметическая	59,9	3,9	62,4	100,0	414,8	111,8	7,5	2,8	13,6	8,7	16,7	1,8
Медиана	59,5	3,1	61,6	84,7	354,4	67,7	6,4	0,6	13,7	8,3	17,0	1,7
Мода	57,0	3,1	62,8	—	—	—	3,6	0,4	7,1	8,3	21,1	1,7
Дисперсия (исправленная)	22,0	10,9	19,9	2266,0	26728,6	27040,6	19,8	29,8	32,4	4,9	11,1	0,4

Окончание табл. 3

Характеристика	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Средне-квадратическое отклонение	4,7	3,3	4,5	47,6	163,5	164,4	4,5	5,5	5,7	2,2	3,3	0,7
Коэф-нт вариации	7,8	83,7	7,2	47,6	39,4	147,2	59,1	192,2	42,0	25,6	20,0	36,7
Коэф-нт асимм.	0,9	4,9	1,1	2,6	1,8	4,7	1,0	2,9	-0,2	1,1	-0,8	3,3
Коэффициент эксцесса	2,2	31,8	2,6	7,4	3,1	24,3	0,7	8,2	-0,9	1,4	1,4	15,7

Наблюдаем большой диапазон изменения показателей по субъектам РФ. Это подтверждает коэффициент вариации, который для всех показателей, кроме уровня занятости населения (X1), доли участия населения в рабочей силе (X3), удельного веса занятых в строительстве (X10) и удельного веса занятых в торговле (X11), говорит о неоднородности данных.

Средние арифметические, медианы и модальные значения демонстрируют расхождения для всех признаков, кроме доли участия в рабочей силе (X3), удельного веса занятых в строительстве (X10) и удельного веса занятых в области информации и связи (X12).

Исходя из данных табл. 3 можно сделать вывод, что все исследуемые показатели отклоняются от нормального распределения: удельный вес занятых в торговле, ремонте автотранспортных средств и мотоциклов (X11) имеет умеренную левостороннюю асимметрию; удельный вес занятых в обрабатывающем производстве (X9) – несущественную.

Остальные показатели имеют правостороннюю асимметрию. В целом, данные табл. 3 выявляют сложность и многообразие исследуемых показателей.

На рис. 2 представлены гистограммы распределения субъектов РФ по значениям некоторых показателей, представленных в табл. 2. Расчеты приведены для 2023 года.

При анализе гистограмм видно, что по каждому рассматриваемому показателю можно выполнить разбиение субъектов на число групп от 3 до 7. В исследовании был использован «метод локтя».

Метод показал, что сумма внутрикластерных расстояний между субъектами по величине исследуемых показателей перестает существенно уменьшаться при числе кластеров  $k=4$ . Поэтому субъекты РФ распределены на 4 кластера. Результаты кластеризации по показателям рынка труда представлены в табл. 4.

Изучим характеристики полученных кластеров. Они приведены в табл. 5.

Таблица 4. Кластеры субъектов РФ по показателям рынка труда в 2023 году

Table 4. Clusters of Russian regions by labor market indicators in 2023

Кластер	Субъекты РФ
1	Тамбовская обл., Респ. Карелия, Респ. Коми, Архангельская обл., Архангельская обл., Респ. Адыгея, Респ. Калмыкия, Астраханская обл., Респ. Дагестан, Респ. Ингушетия, Кабардино-Балкарская Респ., Карачаево-Черкесская Респ., Респ. Северная Осетия – Алания, Чеченская Респ., Респ. Мордовия, Курганская обл., Респ. Алтай, Республика Тыва, Кемеровская обл., Респ. Бурятия, Забайкальский край, Еврейская авт. обл.
2	Ненецкий авт. округ, Ямало-Ненецкий авт. округ, Чукотский авт. округ
3	г. Москва, Мурманская обл., г. Санкт-Петербург, Тюменская обл., Ханты-Мансийский авт. округ – Югра, Респ. Саха (Якутия), Камчатский край, Магаданская обл., Сахалинская обл.
4	Белгородская обл., Брянская обл., Владимирская обл., Воронежская обл., Ивановская обл., Калужская обл., Костромская обл., Курская обл., Липецкая обл., Московская обл., Орловская обл., Рязанская обл., Смоленская обл., Тверская обл., Тульская обл., Ярославская обл., Вологодская обл., Калининградская обл., Ленинградская обл., Новгородская обл., Псковская обл., Респ. Крым, Краснодарский край, Волгоградская обл., Ростовская обл., г. Севастополь, Ставропольский край, Респ. Башкортостан, Республика Марий Эл, Республика Татарстан, Удмуртская Респ., Чувашская Респ., Пермский край, Кировская обл., Нижегородская обл., Оренбургская обл., Пензенская обл., Самарская обл., Саратовская обл., Ульяновская обл., Свердловская обл., Тюменская обл., Челябинская обл., Респ. Хакасия, Алтайский край, Красноярский край, Иркутская обл., Новосибирская обл., Омская обл., Томская обл., Приморский край, Хабаровский край, Амурская обл.

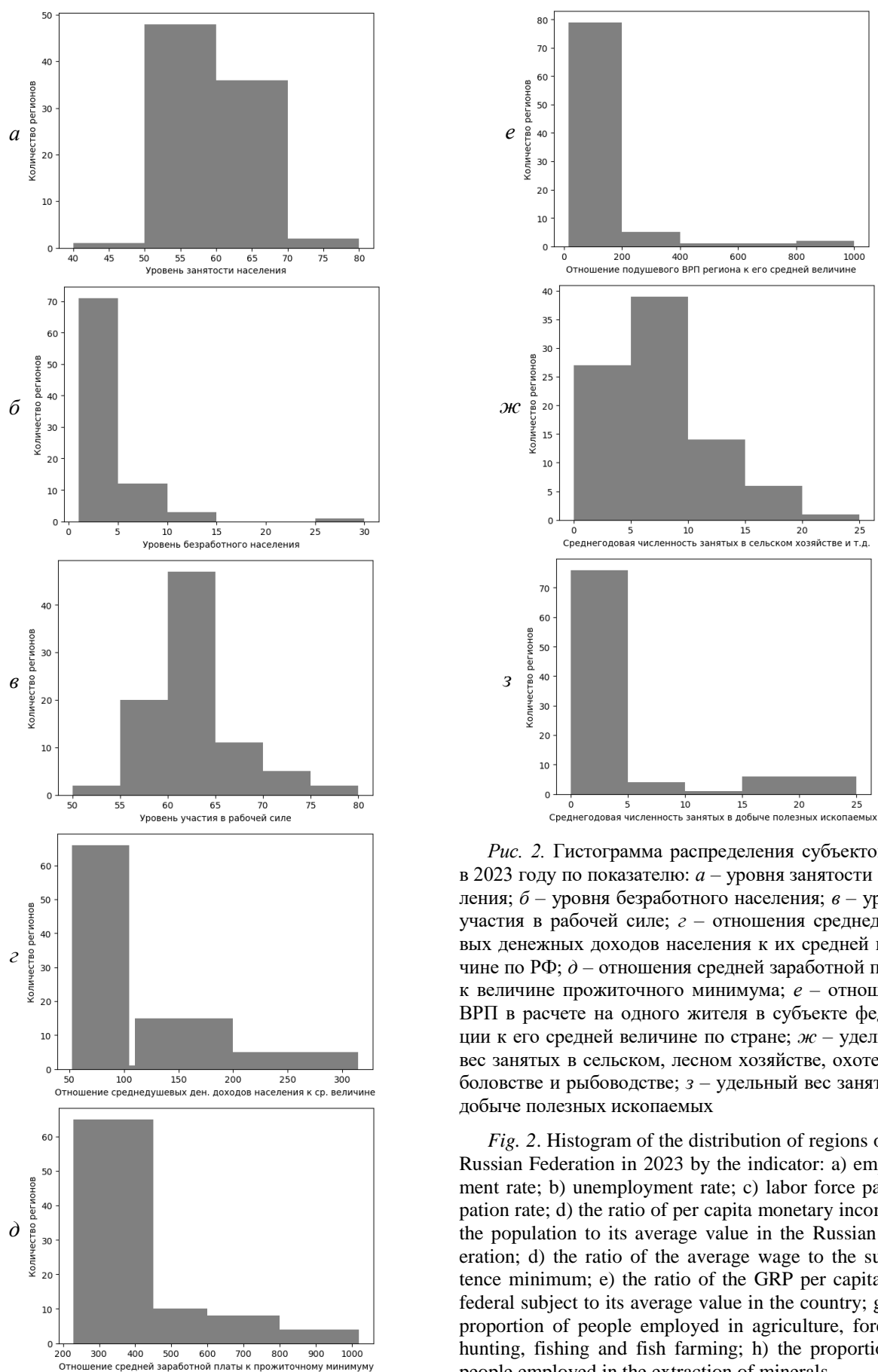


Рис. 2. Гистограмма распределения субъектов РФ в 2023 году по показателю: а – уровня занятости населения; б – уровня безработного населения; в – уровня участия в рабочей силе; г – отношения среднедушевых денежных доходов населения к их средней величине по РФ; д – отношения средней заработной платы к величине прожиточного минимума; е – отношение ВРП в расчете на одного жителя в субъекте федерации к его средней величине по стране; ж – удельный вес занятых в сельском, лесном хозяйстве, охоте, рыболовстве и рыбоводстве; з – удельный вес занятых в добыче полезных ископаемых

Fig. 2. Histogram of the distribution of regions of the Russian Federation in 2023 by the indicator: a) employment rate; b) unemployment rate; c) labor force participation rate; d) the ratio of per capita monetary income of the population to its average value in the Russian Federation; d) the ratio of the average wage to the subsistence minimum; e) the ratio of the GRP per capita in a federal subject to its average value in the country; g) the proportion of people employed in agriculture, forestry, hunting, fishing and fish farming; h) the proportion of people employed in the extraction of minerals



Таблица 5. Средние значения показателей рынка труда в кластерах субъектов РФ в 2023 году

Table 5. Average values of labor market indicators in clusters of regions of the Russian Federation in 2023

Кластер	Количество субъектов, входящих в кластер	Среднее значение показателя в кластере											
		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	22	55,4	7,2	59,9	78,5	343,9	55,5	11,1	1,9	10,9	7,9	15,5	1,4
2	3	73,1	3,2	75,4	281,8	900,9	818,2	3,1	23,4	2,3	10,3	6,8	1,5
3	9	66,1	2,8	68,0	170,9	728,2	263,9	3,6	8,9	7,5	9,7	13,8	2,4
4	53	60,0	2,8	61,8	86,5	363,5	69,3	6,9	1,0	16,4	10,1	18,2	1,8
Среднее значение показателя по РФ		60,8	3,2	62,8	100,0	414,8	111,8	6,0	1,7	14,2	9,3	18,4	2,4

В первый кластер входит 22 субъекта. Здесь мы наблюдаем низкий уровень занятости, который составляет всего 55,4 % и относительно высокий – безработицы (7,2 % населения).

Также в этих субъектах зафиксировано наименьшее среди кластеров отношение среднелюдных денежных доходов населения к их средней величине по РФ (78,5 %), отношение средней заработной платы к величине прожиточного минимума (343,9 %) и наименьшее значение удельного ВРП (55,5 %). В представленных 22 субъектах федерации, входящих в первый кластер, значительная часть населения занимается торговлей, сельским хозяйством, обрабатывающим производством и строительством.

Второй кластер (3 субъекта) характеризуется высоким показателем занятости населения (73,1 %), высокой долей участия в рабочей силе (75,4 %), максимальными среди кластеров отношением среднелюдных денежных доходов населения к их средней величине по РФ (281,8 %) и отношением удельного ВРП в субъекте к его средней величине (818,2%). Почти четверть населения этих субъектов (23,4 %) занята в добывающей отрасли.

Третий кластер (9 субъектов) характеризуется более высокими показателями занятости населения (66,1 %) и доли участия в рабочей силе (68,0 %) по сравнению со средними значениями в РФ. Уровень безработицы минимальный среди кластеров (2,8 %).

Помимо этого, зафиксирован наибольший удельный вес занятых в области информации и связи. Значительная часть населения третьего кластера занята в отраслях торговли и строительства.

Четвертый кластер является наиболее обширным: он включает остальные 53 субъекта федерации.

Его можно охарактеризовать низкими показателями уровня занятости населения, доли участия в рабочей силе, отношения среднелюдных денежных доходов населения к их средней величине по стране, отношения средней заработной платы к величине прожиточного минимума и отношения ВРП в расчете на одного жителя в субъекте федерации к его средней величине по стране по сравнению со средним значением в РФ.

В этих субъектах уровень безработицы составляет всего 2,8 %. Значительная часть населения входящих в четвертый кластер 53 субъектов федерации занимается торговлей и обрабатывающим производством.

На последнем этапе обработки данных формируется интерактивная карта. Результат разработки системы отображения показателей рынка труда на интерактивной карте РФ представлен на рис. 3, а и б.

На интерактивной карте при наведении курсора на субъект выводится информация о названии субъекта РФ, его принадлежности к федеральному округу, величине уровня занятости населения в 2023 году.

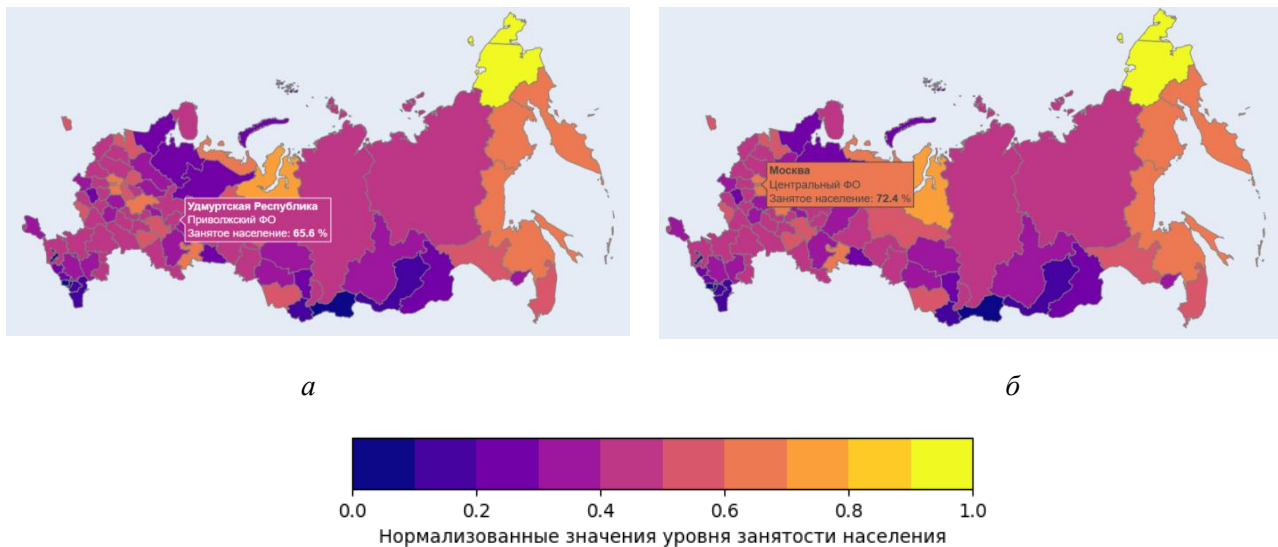


Рис. 3. Интерактивная карта распределения регионов РФ по уровню занятости населения в 2023 году (выводится информация о регионе в зависимости от выбора субъекта федерации): а – Удмуртская Республика; б – г. Москва

Fig. 3. Interactive map of the distribution of regions of the Russian Federation by the level of employment of the population in 2023 (information about the region is displayed depending on the choice of the subject of the federation): a) Udmurt Republic; b) Moscow

### Заключение

Применение информационных технологий обработки данных предоставляет возможность для анализа текущей ситуации на рынке труда и прогноза ее изменений. Обработка статистических данных включает в себя методы для анализа, интерпретации и визуализации информации. В работе рассмотрены составляющие анализа данных, а именно описательная статистика, кластеризация и геоаналитика.

Выполненная кластеризация субъектов РФ по характеристикам рынка труда по статистическим данным за 2023 год определила четыре группы субъектов со схожими показателями рынка труда по уровню занятости и безработицы населения, уровню рабочей силы, отношению среднедушевых денежных доходов населения к их средней величине по стране, отношению среднемесячной заработной платы к прожиточному минимуму, отношению валового регионального продукта в расчете на одного жителя в субъекте федерации к его средней величине по стране, а также по структуре занятого населения в основных видах экономической деятельности.

В первый кластер входит 22 субъекта РФ. Этот кластер характеризуется низким уровнем занятости (55,4%) и высоким уровнем безработицы (7,2%). Здесь зафиксировано наименьшее среди кластеров отношение среднедушевых денежных доходов населения к их средней величине по РФ и наименьшее значение удельного ВРП. В пред-

ставленных 22 субъектах, входящих в первый кластер, значительная часть населения занимается торговлей, сельским хозяйством, обрабатывающим производством и строительством.

Второй кластер (3 субъекта) характеризуется высоким показателем занятости населения (73,1%), высокой долей участия в рабочей силе (75,4%), максимальным среди кластеров отношением среднедушевых денежных доходов населения к их средней величине по РФ (281,8 %) и отношением удельного ВРП в субъекте к его средней величине по стране (818,2 %). Почти четверть населения этих субъектов занята в добывающей отрасли.

Третий кластер (9 субъектов) характеризуется более высокими показателями занятости населения (66,1 %) и долей участия в рабочей силе (68,0 %) по сравнению со средними значениями в РФ. Уровень безработицы минимальный среди кластеров (2,8 %). Помимо этого, зафиксирован наибольший удельный вес занятых в области информации и связи. Значительная часть населения третьего кластера занята в отраслях торговли и строительства.

Четвертый кластер является наиболее обширным: он включает остальные 53 субъекта федерации. Данный кластер можно охарактеризовать низкими показателями уровня занятости населения (60,0 %), доли участия в рабочей силе (61,8 %), отношения среднедушевых денежных доходов населения к их средней величине по стране (86,5 %), отношения средней заработной

платы к величине прожиточного минимума (363,5 %) и отношения ВРП в расчете на одного жителя в субъекте федерации к его средней величине по стране (69,3 %) по сравнению со средним значением в РФ.

В этих субъектах уровень безработицы составляет всего 2,8 %. Значительная часть населения входящих в этот кластер субъектов занимается торговлей и обрабатывающим производством.

Сделан вывод, что представленный в работе комплексный анализ может помочь в получении оценок состояния региональных рынков труда. Использование современных методов обработки данных позволяет получить объективные и детализированные оценки показателей рынка труда в РФ, что способствует принятию обоснованных управленческих решений.

Распределение субъектов РФ по кластерам на основании показателей рынка труда в них позволяет выявить общие тенденции и специфические характеристики, присущие группам субъектов со схожими проблемами и возможностями.

#### Библиографические ссылки

1. Бирюков А. А., Терещенко И. А. Перспективы применения технологии обработки естественного языка в рамках цифровой трансформации экономики РФ и построения цифрового общества // Информационное общество. 2023. № 2. С. 105–120. DOI: 10.52605/16059921\_2023\_02\_105.
2. Ясашин В. А., Болотоков А. С. Модель управления организацией: от статистики к информационным технологиям. Создание единого сквозного процесса обработки и контроля всех данных // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 6 (81). С. 267–270.
3. Менцев А. У., Айсумов Т. Г., Амирова Э. Ф. Методы и технологии сбора и анализа данных в цифровой экономике // Экономика: вчера, сегодня, завтра. 2022. Т. 12, № 11-1. С. 282–288. DOI: 10.34670/AR.2022.36.84.039.
4. Иванов М. Ю., Сыготица М. В. Технологии интеллектуального анализа данных в решении экономических задач // Baikal Research Journal. 2022. Т. 13, № 2. DOI: 10.17150/2411-6262.2022.13(2).27.
5. Курытков В. А., Конюшевская К. О. Прогнозирование рынка труда региона методами системной динамики // Вестник Российского экономического университета имени Г. В. Плеханова. 2012. № 9 (51). С. 94–100.
6. Речинский А. В., Семенов К. К., Сушников В. А., Черненко Л. В. Инструментальные средства систем имитационного моделирования // Известия Тульского государственного университета. Технические науки. 2024. № 4. С. 122–129. DOI: 10.24412/2071-6168-2024-4-122-123.
7. Ванкевич Е. В., Калиновская И. Н. Использование больших данных при анализе рынка труда: теоретические подходы и методические инструменты // Цифровая трансформация. 2024. Т. 30, № 4. С. 23–32. DOI: 10.35596/1729-7648-2024-30-4-23-32.
8. Koch M., Krohmer D., Naab M., Rost D., Trapp M. A matter of definition: Criteria for digital ecosystems // Digital Business. 2022. Vol. 2. 100027. DOI: 10.1016/j.digbus.2022.100027.
9. Камарова Т. А., Маркова Т. Л., Тонких Н. В. Влияние цифровизации занятости на совмещение профессиональных и семейных обязанностей: субъективные оценки россиян // Экономические и социальные перемены: факты, тенденции, прогноз. 2023. Т. 16, № 6. С. 252–269. DOI: 10.15838/esc.2023.6.90.15.
10. Коржова О. С., Стуken Т. Ю., Лапина Т. А., Коржов Е. В. Влияние цифровизации бизнес-процессов предприятий на производительность труда персонала и занятость населения в Российской Федерации // Экономика труда. 2023. Т. 10, № 1. С. 171–180. DOI: 10.18334/et.10.1.116939.
11. Кондратьев Р. А. Влияние цифровых навыков на конкурентоспособность рабочей силы на региональных рынках труда в современных условиях // Вестник Российского нового университета. Серия: Человек и общество. 2025. № 1. С. 14–22. DOI: 10.18137/RNU.V9276.25.01.P.014.
12. Akhatov A. R., Renavikar A., Rashidov A. E. O., Nazarov F. M. Optimization of the number of databases in the big data processing // Проблемы информатики. 2023. № 1 (58). С. 33–47. DOI: 10.24412/2073-0667-2023-1-33-47.
13. Вавилова Д. Д., Кетова К. В. Информационно-аналитическая система анализа региональных социально-экономических процессов на основе комплексного использования динамических моделей различных типов // Прикладная информатика. 2023. Т. 18, № 4 (106). С. 97–110. DOI: 10.37791/2687-0649-2023-18-4-97-110.
14. Вавилова Д. Д., Давлетишина К. М. Data-анализ уровня занятости населения в разрезе по субъектам Российской Федерации // Цифровая и отраслевая экономика. 2023. № 3 (31). С. 93–101.
15. Созинова А. А., Ряттель А. В., Савельева Н. К. Кластерный подход к оценке показателей рынка труда: кросс-региональное сравнение // Экономика труда. 2022. Т. 9, № 10. С. 1509–1526. DOI: 10.18334/et.9.10.116372.
16. Баймуратов И. Р. Методы автоматизации машинного обучения. СПб. : Университет ИТМО. 2020. 40 с.
17. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. М. : ЮНИТИ, 2008. 1005 с.

#### References

1. Biryukov A.A., Tereshchenko I.A. [Prospects for the use of natural language processing technology in the framework of the digital transformation of the Russian economy and the construction of a digital society]. In-

formacionnoe obshchestvo. 2023. No. 2. Pp. 105-120 (in Russ.). DOI: 10.52605/16059921\_2023\_02\_105.

2. Yasashin V.A., Bolotkov A.S. [The management model of an organization: from statistics to information technologies. creation of a single end-to-end process for processing and controlling all data]. *Informacionno-ekonomicheskie aspekty standartizatsii i tekhnicheskogo regulirovaniya*. 2024. No. 6 (81). Pp. 267-270 (in Russ.).

3. Mentsiev A.U., Aigumov T.G., Amirova E.F. [Methods and technologies of data collection and analysis in the digital economy]. *Ekonomika: vchera, segodnya, zavtra*. 2022. Vol. 12, no. 11-1. Pp. 282-288 (in Russ.). DOI: 10.34670/AR.2022.36.84.039.

4. Ivanov M.Yu., Sygotina M.V. [Technologies of data mining in solving economic problems]. *Baikal Research Journal*. 2022. Vol. 13. No. 2. (in Russ.). DOI: 10.17150/2411-6262.2022.13(2).27.

5. Kuryatkov V.A., Konyushevskaya K.O. [Forecasting the regional labor market using methods system dynamics]. *Vestnik Rossijskogo ekonomicheskogo universiteta imeni G. V. Plekhanova*. 2012. No. 9. Pp. 94-100 (in Russ.).

6. Rechinsky A.V., Semenov K.K., Sushnikov V.A., Chernenkaya L.V., Chernenkiy A.V. [Instrumental means of simulation modeling systems]. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*. 2024. No. 4. Pp. 122-129 (in Russ.). DOI: 10.24412/2071-6168-2024-4-122-123.

7. Vankevich E.V., Kalinovskaya I.N. [The use of big data in labor market analysis: theoretical approaches and methodological tools]. *Cifrovaya transformatsiya*. 2024. Vol. 30, no. 4. Pp. 23-32 (in Russ.). DOI: 10.35596/1729-7648-2024-30-4-23-32.

8. Koch M., Krohmer D., Naab M., Rost D., Trapp M. [A matter of definition: Criteria for digital ecosystems]. *Digital Business*. 2022. Vol. 2. 100027. DOI: 10.1016/j.digbus.2022.100027.

9. Kamarova T.A., Markova T.L., Tonkikh N.V. [The impact of digitalization of employment on the combination of professional and family responsibilities: subjective estimates of Russians]. *Ekonomicheskie i social'nye peremeny: fakty, tendencii, prognoz*. 2023. Vol.

16, no. 6. Pp. 252-269 (in Russ.). DOI: 10.15838/esc.2023.6.90.15.

10. Korzhova O.S., Stuken T.Yu., Lapina T.A., Korzhov E.V. [The impact of digitalization of business processes of enterprises on staff productivity and employment in the Russian Federation]. *Ekonomika truda*. 2023. Vol. 10, no. 1. Pp. 171-180 (in Russ.). DOI: 10.18334/et.10.1.116939.

11. Kondratiev R.A. [The impact of digital skills on the competitiveness of the workforce in regional labor markets in modern conditions]. *Vestnik Rossijskogo novogo universiteta. Seriya: CHelovek i obshchestvo*. 2025. No. 1. Pp. 14-22 (in Russ.). DOI: 10.18137/RNU.V9276.25.01.P.014.

12. Akhatov A.R., Renavikar A., Rashidov A.E.O., Nazarov F.M. [Optimization of the number of databases in the big data processing]. *Problemy informatiki*. 2023. No. 1. Pp. 33-47. DOI: 10.24412/2073-0667-2023-1-33-47.

13. Vavilova D.D., Ketova K.V. [Information and analytical system for analyzing regional socio-economic processes based on the integrated use of dynamic models of various types]. *Prikladnaya informatika*. 2023. Vol. 18, no. 4. Pp. 97-110. (in Russ.). DOI: 10.37791/2687-0649-2023-18-4-97-110.

14. Vavilova D.D., Davletshina K.M. [Data analysis of the level of employment in the context of the subjects of the Russian Federation]. *Cifrovaya i otraslevaya ekonomika*. 2023. No. 3. Pp. 93-101 (in Russ.).

15. Sozinova A.A., Rattel A.V., Savelyeva N.K. [Cluster approach to assessing labor market indicators: a cross-regional comparison]. *Ekonomika truda*. 2022. Vol. 9, no. 10. Pp. 1509-1526 (in Russ.). DOI: 10.18334/et.9.10.116372.

16. Baymuratov I.R. *Metody avtomatizatsii mashin-nogo obucheniya* [Methods of automating machine learning]. St. Petersburg: ITMO University. 2020. 40 p. (in Russ.).

17. Ayvazyan S.A., Mkhitarian V.S. *Prikladnaya statistika i osnovy ekonometriki* [Applied statistics and fundamentals of econometrics]. Moscow, UNITY. 2008. 1005 p. (in Russ.).

\* \* \*

### Information Technologies for Data Processing to Analyze Labor Market Indicators by the Russian Federation Regions

D. D. Vavilova, PhD in Engineering, Associate Professor, Kalashnikov Izhevsk State Technical University, Izhevsk, Russia

K. V. Ketova, Doctor of Physics and Mathematics, Professor, Kalashnikov Izhevsk State Technical University, Izhevsk, Russia

K. M. Davletshina, Student, Kalashnikov Izhevsk State Technical University, Izhevsk, Russia

The article is devoted to the study and classification of information technologies for statistical data processing being applied to labor market analysis. The methods that contribute to understanding the employment structure and other aspects of the labor market in the Russian Federation are considered. The emphasis is placed on the use of descriptive statistics, clusterization and geo-analytics of labor market indicators in the context of the regions of the Russian Federation. The result of the research is the analysis of the labor market characteristics of the regions, statistical and correlation data analysis, clustering of the Russian Federation regions according to labor market indicators, as well

as the creation of an interactive map based on the employment level. The clustering identified groups of regions with similar labor market indicators in terms of employment and unemployment, labor force level, the ratio of per capita monetary incomes of the population to their national average, the ratio of average monthly wages to the subsistence minimum, the ratio of the specific gross regional product in the subject to its national average, as well as the structure of the employed population in the main types of economic activity. It was found that in 2023, 4 clusters are being formed in the Russian labor market with regions whose indicators within each cluster have similar values. The first cluster includes 22 regions of the Russian Federation with low employment (55.4%) and high unemployment (7.2%) of the population. The second cluster includes 3 regions and is characterized by a high employment rate (73.1%), a high share of labor force participation (75.4%), and the highest ratio of per capita monetary incomes of the population to their average value in the Russian Federation (281.8%). A significant part of the population in these regions is engaged in mining (23.4%). The third cluster includes 9 regions of the Russian Federation with higher employment rates (66.1%) and labor force participation (68.0%) compared to the average in the Russian Federation, as well as the lowest unemployment rate among the clusters. The fourth cluster is the most extensive: it includes the remaining 53 subjects of the Russian Federation. It can be characterized by low values of the considered indicators. The regions of this cluster also have the lowest unemployment rate (2.8%).

**Keywords:** information technologies, data analysis, clusterization, labor market, region.

Получено: 27.03.25

#### Образец цитирования

Вавилова Д. Д., Кетова К. В., Давлетишина К. М. Информационные технологии обработки данных для анализа показателей рынка труда в разрезе по субъектам Российской Федерации // Интеллектуальные системы в производстве. 2025. Т. 23, № 2. С. 80–92. DOI: 10.22213/2410-9304-2025-2-80-92.

#### For Citation

Vavilova D.D., Ketova K.V., Davletshina K.M. [Information Technologies for Data Processing to Analyze Labor-Market Indicators by the Russian Federation Regions]. *Intellectual'nye sistemy v proizvodstve*. 2025, vol. 23, no. 2, pp. 80-92. DOI: 10.22213/2410-9304-2025-2-80-92.