

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский экономический университет имени Г.В. Плеханова»

Высшая школа кибертехнологий, математики и статистики

Направление Бизнес-информатика

Кафедра Информатики

КУРСОВОЙ ПРОЕКТ

по дисциплине «Информационный менеджмент»

на тему: «Разработка концепции распределенной информационной системы
построения модели угроз и управления гражданскими рисками»

Выполнила
обучающаяся группы
15.11Д-БИЦТ09/216
очной формы обучения
Высшей школы кибертехнологий,
математики и статистики
Крабу Кира Сергеевна

Научный руководитель:
Брежнев А.В., к.н.,
доцент кафедры информатики

Москва 2025

Оглавление

Введение	3
Основная часть	5
Глава 1. Система управления рисками	5
1.1. Общие сведения о системе управления рисками.....	5
1.2. Существующие системы управления рисками	6
1.3. Правовая основа	7
Глава 2. Концепция распределенной информационной системы управления рисками	9
2.1. Основные понятия и принцип работы	9
2.2. Модели угроз и рисков	11
2.3. Математические модели	12
2.4. Источники данных	14
2.5. Планы реагирования на риски	15
Глава 3. Техническая реализация проекта	17
3.1. Описание реализуемой части	17
3.2. Выбор данных для реализации	18
3.3. Анализ данных	23
3.4. Разработка ВІ-системы.....	28
3.5. Оценка результатов	30
Заключение	32
Список использованных источников.....	34
Приложение 1	38
Приложение 2	39
Приложение 3	40
Приложение 4	41

Введение

Современный мир – постоянно изменяющееся место, динамично развивающаяся система, которая привносит в жизнь людей много нового. В опыте человека появляются ранее не изведанные или редко появляющиеся ситуации, из-за этого растет неопределенность. У людей появляется потребность в формализации того, что они уже знают и с чем могут бороться. Управление рисками играет важную роль как в этой деятельности, так и в обеспечении стабильного развития любого современного государства. Система управления рисками решает задачу агрегирования возможных угроз, тем самым создавая «базу знаний» в выбранной сфере, идентифицируя и оценивая возможные риски. Это позволяет лучше их ранжировать, минимизировать потери и оптимизировать использование ресурсов на ликвидацию негативных последствий. Такая система актуальна для гражданской сферы общества, где угрозы могут приводить к социально-экономическим и политическим последствиям и быть максимально ощутимыми для граждан и государства.

В нынешней ситуации, информационный шум влияет на повседневную жизнь всех граждан и мешает концентрации на конкретных проблемах. Это касается и государственного аппарата Российской Федерации. Внедрение такой системы может иметь положительную динамику влияния на ситуацию, когда фокус внимания максимально рассеян между внутренней и внешней политикой.

Целью данного курсового проекта является создание концепции распределенной информационной системы построения модели угроз и управления рисками в гражданской сфере.

К задачам рассматриваемой работы можно отнести:

- Исследование уже существующих примеров СУР;
- Анализ математической базы;
- Реализация простого проекта описываемой системы.

Предмет исследования – система управления рисками и построение модели угроз. Объект – распределенная информационная система построения модели угроз и управления рисками в гражданской сфере.

Методы исследования в данном курсовом проекте:

- Изучение документов;
- Анализ литературы;
- Сравнительный анализ;
- Описание систем;
- Систематизация;
- Моделирование;
- Тестирование.

В первой главе будут описаны теоретические основы для реализации проекта: описание концепции систем управления рисками, описание примеров, которые уже задействованы в Российской Федерации, правовые основы для их деятельности.

Во второй главе будет подробно описана реализуемая в данном проекте концепция: ее архитектура, математическая база, используемые технологии, модули и источники данных.

В заключительной части работы, в третьей главе, будет приведено описание работ по созданию распределенной информационной системы построения модели угроз и управления рисками: выбранные инструменты и шаги по реализации. В конце проекта будут сделаны выводы.

Основная часть

Глава 1. Система управления рисками

1.1. Общие сведения о системе управления рисками

Система управления рисками (СУР) – это такой подход к деятельности организации или, в данном случае, страны, который включает в себя методы, процессы, политики по идентификации угроз и снижению их влияния, минимизацию убытков. Понятие риска [19] существует во многих сферах жизни человека – это возможность опасности, неудачи [13].

Угроза в контексте гражданской сферы может определяться как потенциальное или реальное событие, которое может нанести вред жизни, здоровью и имуществу граждан, а также нарушить общественный порядок и функционирование общественных институтов. Гражданские угрозы могут проявляться в различных формах, включая природные катастрофы, техногенные аварии, террористические атаки, эпидемии, и другие факторы, способные вызвать негативные последствия для населения и окружающей среды.

Можно разделить подходы к управлению рисками [8]:

- Активный подход предполагает максимальное задействование ресурсов и средств для управления и минимизации потерь;
- Адаптивный подход основан на учете сложившихся условий;
- Консервативный подход к системе управления рисками подразумевает действия уже после наступления последствий и направлен на локализацию уже осуществленного ущерба.

По организационной структуре СУР могут быть спроектированы как:

- Централизованная модель, которая подразумевает, что все процессы – анализ, оценка и управление рисками – сосредоточены в одном структурном подразделении;

- Распределенная модель предполагает наличие небольшого отдела, занимающегося мониторингом рисков, в то время как функции управления распределяются между другими [8].

Для реализации данного проекта подразумевается распределенная модель СУР и активным или адаптивным подходом к работе.

1.2. Существующие системы управления рисками

Для начала, предлагается рассмотреть уже введенные в эксплуатацию системы управления рисками в некоторых сферах деятельности Российской Федерации [23]. На таможенной службе Российской Федерации СУР используется с 2002 года. Таможенные органы РФ в соответствии таможенным кодексом Евразийского экономического союза используют [1] СУР для выбора объектов таможенного контроля и мер по минимизации рисков. Впоследствии был также утвержден документ о «Стратегии и тактике применения СУР, а также процедуры сбора, обработки информации, проведения анализа и оценки рисков, разработки и реализации мер по их управлению (за исключением рисков в области ветеринарии, санитарно-эпидемиологического надзора и обеспечения карантина растений)» [5]. Благодаря этому документу сформировались унифицированные подходы к управлению, внедрению и совершенствованию системы.

Под управлением рисками на таможне подразумевается деятельность, которая контролирует соблюдение международных договоров, «актов в области таможенного регулирования, а также законодательства государств-членов» [26].

Основными целями СУР на таможне РФ выделяются [8]:

- Повышение эффективности таможенного контроля;
- Повышенное внимание областям риска и эффективное использование ресурсов таможни;

- Создание специальных упрощенных и ускоренных условий для перемещения товаров, для которых риск минимален или не выявлен.

В Ханты-Мансийском автономном округе есть своя система управления рисками, которая в 2022 году учитывала «причинение вреда (ущерба) охраняемым законом ценностям при осуществлении регионального государственного контроля (надзора)». Платформа объединяла риски в таких областях как защита населения и территорий от ЧС, контроль в транспорте, социальное обслуживание, архивные дела и так далее [18]. В 2023 году в обзорном документе по данной была опубликована частота проверки угроз на предмет возникновения риска и его степени:

- Если риск значительный – ежегодно;
- Если риск средний – не реже 1 раза в 2 года;
- Если риск умеренный – не реже 1 раза в 3 года;
- Если риск низкий – не реже 1 раза в 5 лет.

Данная информация может пригодится в дальнейшем.

1.3. Правовая основа

В основе работы с рисками на федеральном уровне лежат государственные документы – например, доктрины. На данный момент доктрин по безопасности в Российской Федерации всего шесть основных, подписанных Президентом РФ [16]: военная, климатическая, продовольственная, информационная, энергетическая и морская. В данных документах находятся системы государственных взглядов на основные цели, принципы, задачи и механизмы в конкретной области. К сожалению, не все сферы государственной и общественной жизни затрагиваются этими документами.

При разработке РИС нужно опираться не только на основные правовые документы, но и на доктрины, национальные программы и проекты, ГОСТы. Одним из основных документов может послужить указ Президента

Российской Федерации «О Стратегии национальной безопасности Российской Федерации» [3], а также «О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года» [2]. В качестве технологической основы можно рассматривать национальный проект «Цифровая экономика» [27] и «Положение о единой межведомственной информационно-статистической системе» (ЕМИСС) [4], используя последнее для примера.

Глава 2. Концепция распределенной информационной системы управления рисками

2.1. Основные понятия и принцип работы

Информационная система – это комплекс программных и аппаратных средств, который предназначен для сбора, хранения, обработки и выдачи информации.

Распределенная информационная система – это набор информационных систем (программного и аппаратного обеспечения) и баз данных, расположенных на нескольких узлах и работающих дистанционно друг от друга. Узлы могут быть и как виртуальными, так и физическими, и взаимодействовать через общую сеть. Такая технология позволяет сократить нагрузки на серверы и обеспечить работы подразделений, расположенных в разных географических точках. Поскольку предлагаемая в данной работе система федерального масштаба, такой подход будет очень актуален.

Модель угроз – это структурированное представление информации или подход к выявлению и анализу потенциальных угроз. Модель угроз помогает понять, какие риски существуют, какова их вероятность и источник возникновения, какие последствия ожидаются. В контексте данной работы также подразумевается, что модель угроз классифицирует данные и предлагает пути минимизации рисков.

Гражданские риски, как уже было описано ранее, это вероятные возникновения негативных событий, которые могут повлиять на жизнь и безопасность граждан Российской Федерации, а также общественных институтов страны.

Как стало понятно, в настоящее время существуют и системы управления рисками, и документы, которые закладывают основу данным системам на федеральном уровне. Однако отсутствие общей платформы может плохо влиять на анализ угроз, а также разрозненность информации о рисках между разными органами власти и ведомствами может препятствовать

комплексному анализу и управлению. Также существующие методы управления рисками часто ориентированы на отдельные виды рисков и не учитывают их взаимосвязи, все из-за той же разрозненности данных.

Для решения этих проблем предлагается создать распределенную информационную систему (РИС) построения модели угроз и управления рисками, которая собирала бы и агрегировала все данные из схожих специализированных систем и из государственных баз данных. Из-за распределённой архитектуры платформа приобретает несколько преимуществ, таких как масштабируемость, повышенная эффективность, отказоустойчивость, оптимальное использование ресурсов [7]. Имея в своем «строении» несколько модулей, можно работать с их иерархией и режимами доступа.

Аппаратное обеспечение описываемой распределенной информационной системы включает в себя несколько десятков или сотен устройств. В первую очередь необходимо обеспечить хранилища данных, т.е. мощные локальные диски, серверы и облака. Далее необходимо иметь персональные компьютеры и мощные ноутбуки с российским ПО и отечественные решения программных роботов для реализации сбора данных и их обработки. В аппаратное обеспечение входят маршрутизаторы, брандмауэры, системы предотвращения вторжений (IPS), балансировщики нагрузки, графические процессоры и аналитические системы, а также системы резервного копирования на магнитных лентах.

По итогу, структура распределенная информационная система построения модели угроз и управления рисками будет состоять из нескольких модулей:

- модуль сбора и обработки данных;
- модуль кластеризации угроз;
- модуль оценки вероятности рисков;

- модуль планов реагирования (он же база знаний и экспертная система);
- модуль мониторинга и управления системой.

2.2. Модели угроз и рисков

Примеры возможных угроз в гражданском контексте представлены в таблице 2.1.

Таблица 2.1. – Гражданские угрозы

Сфера	Пример угрозы
Природные катастрофы	Землетрясения
	Наводнения
	Ураганы и другие стихийные бедствия
Техногенные катастрофы	Аварии на производственных объектах
	Выбросы опасных веществ
	Кибератаки на критически важные инфраструктуры (например, энергосистемы, транспортные сети)
Социальные и политические риски	Массовые беспорядки
	Терроризм
	Преступность
	Социальная напряженность
	Коррупция
	Безработица
	Миграция
Здравоохранение	Нехватка медицинских работников
	Эпидемии и пандемии
	Нехватка лекарственных средств
	Увеличение числа пожилых людей

2.3. Математические модели

После того, как система агрегировала все нужные данные, они подвергаются кластеризации – методу машинного обучения, который группирует их по степени схожести. Это можно реализовать с помощью нескольких способов работы [9, 10]. Например, с иерархической кластеризацией, которая создает дендрограмму, наглядно показывая, как кластеры данных связаны друг с другом. Алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise) считается эффективным для выявления аномалий и редких угроз, поскольку он распределяет объекты по кластерам на основе плотности точек. Смешанные модели кластеризации или же гауссовские смешанные модели (Gaussian Mixture Models) подходят к данным как к комбинации нескольких нормальных распределений, что помогает учитывать сложную структуру данных и неопределенность. Также в РИС необходимо использовать меру TF-IDF (Term Frequency-Inverse Document Frequency) [15], которая позволяет находить похожие риски, угрозы, основываясь на их описании, превращая текст в векторы. В дальнейшем, алгоритмы случайного леса (Random Forest) или градиентного бустинга (Gradient boosting) могут быть использованы для классификации угроз на основе признаков, которые уже были получены в результате кластеризации.

Поскольку кластеризация – это метод машинного обучения без учителя, заранее предвидеть точное количество кластеров является задачей сложной. В качестве примера схожих характеристик, по которым будут отсортированы угрозы, можно лишь представить:

- вид угрозы;
- источник возникновения;
- степень серьезности;
- степень вероятности наступления последствий;
- географический уровень и т.д.

Поскольку РИС агрегирует многие угрозы и риски, нужно учитывать их взаимосвязь. Для того, чтобы определять вероятность возникновения риска, необходимо использовать байесовскую модель (формулы (1) и (2)):

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}, \quad (1)$$

$$P(B|A_1A_2A_3) = \frac{P(B)P(A_1|B)P(A_2|B)P(A_3|B)}{P(A_1A_2A_3)}, \quad (2)$$

где:

- $P(B|A)$ – условная вероятность события В, если произошло событие А;
- $P(A|B)$ – условная вероятность события А, если произошло событие В;
- $P(B)$ – вероятность события В;
- $P(A)$ – вероятность события А.

Если A_1, A_2, A_3 взаимно независимы, то формулу можно расширить.

Помимо того, что подход к расчету вероятности в байесовской статистике [12, 14, 17] включает в себя учет взаимосвязей, она также считывает изменения, которые будут происходить при добавлении новых данных.

Внутри кластеров высчитывание такой вероятности риска необходимо, поскольку там находятся похожие угрозы. Мета-анализ также будет важной частью работы системы. Его можно производить с помощью машинного и глубокого обучения, используя регрессионные модели и нейросети, решающие задачу прогнозирования. Для того, чтобы это могло дать свои плоды, необходимо создать базу знаний, наполняя ее историческими данными об угрозах, рисках и их последствиях.

Поскольку один из принципов работы РИС – взаимосвязанный подход, можно также рассмотреть методы теории игр – математической дисциплины, которая рассматривает взаимодействие «игроков» – в данном случае это могут быть как риски, так и планы реагирования на них [11]. Анализ стратегий поведения одного «участника» может помочь понять, как на него повлияли

действия другого. Также с помощью теории игр можно моделировать, тем самым тестировать различные сценарии рисков и планов реагирования. Это своеобразная эволюционная игра, которая применяется для анализа того, как стратегии управления рисками могут изменяться и адаптироваться со временем в ответ на изменения внешней среды или поведения других участников.

2.4. Источники данных

Как уже было сказано, РИС может объединять уже существующие системы управления рисками, а также обеспечить сбор и интеграцию данных из различных государственных баз данных, таких как: Росстат, Минтруд, банк России, ЕМИСС, ГИС ЖКХ, госавтоинспекция, МВД, Минкульт, Роскомнадзор, ФНС, Минобрнауки, различная правовая информация. Для автоматизации этой работы можно использовать программных роботов, задавая им сценарии заполнения базы знаний РИС. Однако сами угрозы должны прописываться или определяться экспертной группой, «запуская» их в систему. Программные роботы могут собирать данные и выделять возможные проблемы, но последнее слово должно быть за человеком. После задачи кластеризации угроз, будет осуществлен системный подход к анализу рисков, который подразумевает учет взаимосвязей между их разными видами. Также рассматриваемая РИС должна прогнозировать риски, используя современные методы машинного и глубокого обучения (нейронные сети).

Для сбора данных также могут быть использованы не только государственные базы данных, но и некоммерческие организации или другие агрегаторы данных, полезных для понимания ситуации в стране. Например, благотворительные фонды могут послужить источником информации о своих подопечных. Также ссылаться можно на собираемые обращения граждан в различных ведомствах.

2.5. Планы реагирования на риски

Распределенная информационная система управления рисками также включает в себя планы реагирования и оценки бюджета на них. План реагирования должен разрабатываться экспертами и учитывать различные национальные программы. Он должен включать в себя:

- оперативные действия;
- ответственные ведомства и их роли;
- ресурсы (люди, оборудование, финансирование);
- временные рамки;
- критерии оценивания эффективности решения.

Изначально можно использовать уже существующие сценарии. Далее можно создавать шаблоны (промпты), на основе которых будут строиться будущие планы. Данные документы должны храниться в базе знаний – цифровом архиве – это является одним из компонентов блока экспертной системы. Для их написания необходимо подключать экспертов из различных сфер, например обязательным будет являться человек, который ответственен за определение ставки рабочего. Возможно, это будет автоматизировано путем подключения внутренних систем ведомств с данными о заработной плате сотрудников.

Можно привести примеры возможных планов реагирования (оперативных действий):

- эвакуация населения;
- обеспечение временного, социального жилья;
- финансовая помощь;
- оперативное устранение последствий;
- информирование населения;
- меры по поддержке бизнеса;
- социальные программы для граждан и т.д.

Бюджет на такие мероприятия должен учитывать заложенный бюджет, количество затраченных ресурсов и их стоимость, в том числе и заработная плата работников ответственных ведомств. За начальные числа можно взять выделенные средства на различные национальные программы – это около 3 миллиардов рублей на 15 программ [6, 20]: демография, здравоохранение, образование, жилье и городская среда, экология, безопасные качественные дороги, производительность труда, наука и университеты, цифровая экономика, культура, предпринимательство, международная кооперация и экспорт, магистральная инфраструктура, туризм и беспилотные авиационные системы.

Глава 3. Техническая реализация проекта

3.1. Описание реализуемой части

Для реализации технической части данного проекта будет описана работа по созданию базы данных и панель с результатами их анализа – дашборд. Представленная далее система будет являться лишь концепцией и черновым вариантом описываемой системы.

Задание для предстоящей работы будет таковым: для начала нужно определиться с данными, на которых предстоит основать «систему». Далее необходимо создать базу данных для хранения информации, ограничиваясь пока что ручным загрузиванием данных, т.е. без автоматизации. Далее необходимо данные проанализировать и выдать прогнозируемые значения показателей на предстоящие годы. После этого подключается BI-система и прогнозы выставляются на дашборде.

Для реализации этих действий будут использованы исключительно бесплатные виды инструментов. Для создания базы данных была выбрана PostgreSQL [24] – свободная СУБД, которая поддерживает сложные структуры данных и имеет множество совместимостей с различными платформами.

Для проведения анализа данных и прогноза будут использованы такие средства как язык программирования Python и облачная среда программирования Google Colab [22]. Python представляет собой достаточно простой в понимании язык, а также он поддерживает написание нейронных сетей, с помощью которых будет осуществляться прогнозирование показателей. Google Colab имеет большое преимущество перед многими другими средами программирования: он позволяет запускать не весь код сразу, а лишь его часть, что также идет на пользу при обучении нейросетей.

В качестве BI-системы будет использована отечественное облачное решение Yandex DataLens [21] – оно поддерживает подключение к PostgreSQL и имеет весь необходимый функционал для построения дашборда.

3.2. Выбор данных для реализации

Источником данных для работы стала федеральная служба государственной статистики – Росстат [25]. Выбор данных был исключительно субъективным процессом, однако можно выявить некоторые критерии:

- Данные должны точно быть из категории «риска», т.е. иметь четкое определение того, что если с показателями что-то произойдет, будет выявлен риск;
- Данные должны быть распределены по субъектам Российской Федерации;
- Данные должны быть полные;
- Показатели должны быть выражены в числовом формате;
- Данные должны иметь явную связь между собой: даже без математического анализа должно быть ясно, что они влияют на него;
- Должен быть выявлен «главный» показатель, вокруг которого будет все построено.

Для начала была выбрана основная «область» данных и «главный» показатель – численность безработных в возрасте от 15 до 72 лет. Такая статистика ведется с 2000 года и в таблице исчисляется в тысячах человек (рисунок 3.1).

К содержанию											
Численность безработных в возрасте 15-72 лет по субъектам Российской Федерации											
(по данным выборочных обследований рабочей силы)											
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Российская Федерация	7699.5	6423.7	5698.3	5933.5	5666.0	5242.0	5250.2	4518.6	4697.0	6283.7	5544.2
Центральный федеральный округ	1516.4	1165.3	1000.2	1002.1	930.5	852.3	814.6	635.6	732.9	1193.0	960.8
Белгородская область	44.9	48.2	58.5	59.8	45.9	42.0	42.7	31.4	29.6	37.4	40.6
Брянская область	88.8	63.8	54.4	47.7	58.4	43.9	44.8	42.5	42.2	68.2	51.9
Владимирская область	95.5	80.4	83.7	81.1	73.3	72.9	86.3	53.7	44.9	66.5	46.6
Воронежская область	118.6	105.7	101.3	91.5	95.7	87.6	63.7	59.4	60.8	100.8	86.8
Ивановская область	61.4	33.9	39.1	36.5	26.9	39.5	23.5	24.1	29.5	58.8	42.2
Калужская область	49.3	33.1	34.6	32.7	34.7	31.1	30.5	27.4	25.9	35.0	37.6
Костромская область	33.3	22.2	18.5	22.0	21.8	18.0	18.2	12.2	18.1	29.7	21.4
Курская область	66.5	64.1	43.8	49.5	44.4	43.7	42.5	29.1	37.3	51.0	47.6
Липецкая область	52.1	38.4	29.7	27.0	25.6	48.3	30.0	16.7	30.7	34.9	27.8
Московская область	283.5	196.9	152.5	159.8	139.9	117.9	110.8	78.8	105.7	187.4	130.5
Орловская область	37.6	33.8	29.7	34.1	26.6	26.3	25.3	22.9	24.1	39.3	35.5
Рязанская область	57.7	63.8	50.2	49.2	34.0	32.0	28.2	22.4	31.0	51.8	48.3
Смоленская область	67.2	53.5	59.5	56.4	48.1	41.4	43.6	36.9	37.2	43.6	40.8

Рисунок 3.1. – Фрагмент таблицы «Численность безработных»

Отталкиваясь от этого, были выбраны еще такие таблицы как:

- Уровень бедности (доля населения, проживающего ниже границы бедности) по субъектам РФ. Статистика ведется с 2018 по 2023 годы в процентах от общей численности населения на выбранный год. Далее в работе – уровень бедности;
- «Среднемесячная начисленная заработная плата наёмных работников в организациях, у индивидуальных предпринимателей и физических лиц» по субъектам РФ с 2015 по 2023 год в рублях (далее – средняя зарплата);
- Граница бедности, он же прожиточный минимум. Данные распределены по субъектам РФ с 2021 по 2023 годы в рублях и по всему населению (в Росстате этот показатель также имеется по отдельным социально-демографическим группам). Далее в работе – граница бедности или прожиточный минимум;
- «Ожидаемая продолжительность предстоящей жизни по субъектам РФ», описанная как процент людей, достигших 60 лет, с 2013 по 2023 годы (далее – продолжительность жизни).

Также для того, чтобы в дальнейшем разнообразить панель данных были выбраны еще 2 таблицы, в которых не разбивки по субъектам РФ:

- Заболеваемость населения алкоголизмом и алкогольными психозами (данные Минздрава России, расчет Росстата) с 2003 по 2023 годы, измеряемая в тысячах людей (далее – алкоголизм);
- Заболеваемость населения наркоманией (данные Минздрава России, расчет Росстата) с 2003 года по 2023, измеряемая в тысячах людей (далее – наркомания).

Также для того, чтобы можно было создать полноценную базу данных и дашборд в дальнейшем были добавлены несколько других таблиц. «Регионы РФ» содержит список названий субъектов Российской Федерации и их официальными кодами (далее название таблицы в работе – регионы). Таблица

«Regions.csv» – набор с названиями субъектов РФ и их полигонами координат для построения дашборда (рисунок 3.2) (далее – координаты субъектов).

```
Архангель [[65.051193, 35.314801], [65.127703, 35.29793], [65.184477, 35.346113], [65.220663, 35.435205], [65.252509, 35.612738], [
53.021353], [80.550733, 53.01071], [80.57999, 53.063697], [80.600841, 52.933767], [80.625795, 52.898554], [80.666723, 53.026964], [8
Алтайский [[53.290641, 77.907506], [53.367469, 77.889413], [53.356466, 77.942015], [53.388527, 77.96952], [53.420347, 77.916511], [
Амурская [[56.865828, 119.818979], [56.903958, 119.88444], [56.908385, 120.020536], [56.929373, 120.092292], [57.02688, 120.2907
Астрахань [[48.270856, 44.962711], [48.291774, 44.98271], [48.359397, 44.991527], [48.358389, 45.02712], [48.425013, 45.013654], [4
Белгород [[50.97531, 35.352225], [50.958115, 35.411367], [50.948591, 35.421259], [50.957812, 35.4357], [50.955282, 35.456132], [50
Брянская [[53.911337, 34.146682], [53.908911, 34.140989], [53.912425, 34.136441], [53.917479, 34.14842], [53.918199, 34.151421], [
Владимир [[56.524305, 38.268605], [56.526572, 38.279686], [56.53144, 38.273602], [56.531879, 38.285962], [56.547254, 38.299358], [
Волгоград [[50.79865, 41.247877], [50.842047, 41.392559], [50.832007, 41.420391], [50.917199, 41.427638], [50.950281, 41.471398], [
Вологодск [[59.085189, 34.721699], [59.097809, 34.779386], [59.127616, 34.803738], [59.127354, 34.835685], [59.215322, 34.867008],
Воронеж [[51.637092, 38.140099], [51.670452, 38.161229], [51.659269, 38.234334], [51.699828, 38.224065], [51.693568, 38.289492],
Москва [[55.733119, 36.890429], [55.73302, 36.889551], [55.73483, 36.888508], [55.733754, 36.890173], [55.734064, 36.893504], [5
Еврейская [[48.881719, 130.626261], [48.89392, 130.628169], [48.912183, 130.654197], [48.916288, 130.651406], [48.926637, 130.678
Забайкаль [[49.97121, 107.754632], [49.991055, 107.797169], [50.015912, 107.766529], [50.032103, 107.795989], [50.153166, 107.734
Ивановск [[56.771178, 39.400991], [56.779082, 39.406708], [56.776914, 39.378985], [56.783836, 39.378737], [56.78637, 39.385517], [
Иркутская [[54.243839, 95.674504], [54.391303, 95.680534], [54.424562, 95.817534], [54.395886, 95.918024], [54.573935, 96.062808],
Кабардин [[43.256431, 42.414161], [43.274642, 42.415043], [43.290057, 42.396782], [43.317201, 42.409306], [43.342876, 42.413682],
Калининг [[54.537929, 19.484768], [54.663893, 19.663573], [54.702118, 19.683266], [54.738499, 19.731087], [54.753372, 19.741611],
Калужская [[54.01633, 33.50772], [54.023626, 33.508793], [54.027057, 33.493605], [54.033671, 33.491109], [54.026855, 33.462102], [5
Камчатск [[58.667208, 163.399313], [58.670613, 163.431727], [58.743192, 163.471348], [58.79141, 163.535046], [58.834151, 163.565
Карацев [[43.655105, 40.702984], [43.660919, 40.709432], [43.67046, 40.707248], [43.67894, 40.716539], [43.685768, 40.712913], [4
Кемерово [[53.359263, 86.764418], [53.359183, 86.789082], [53.355263, 86.793794], [53.353033, 86.799283], [53.352244, 86.803997],
Кировская [[60.249615, 46.277807], [60.240878, 46.348748], [60.24682, 46.384849], [60.263913, 46.390479], [60.269775, 46.416512], [
Костром [[57.4385, 40.424032], [57.446549, 40.428638], [57.458128, 40.417288], [57.474298, 40.428261], [57.484219, 40.409266], [5
```

Рисунок 3.2. – Фрагмент полигона координат субъекта

Для понимания, какую базу данных следует создать, была нарисована схема данных «Звездочка» (рисунок 3.3). Благодаря этому, можно понять, какую примерную схему будут иметь данные. Таблицы с данными по алкоголизму и наркомании не подгружаются сюда, так как не имеют разбивки по субъектам.

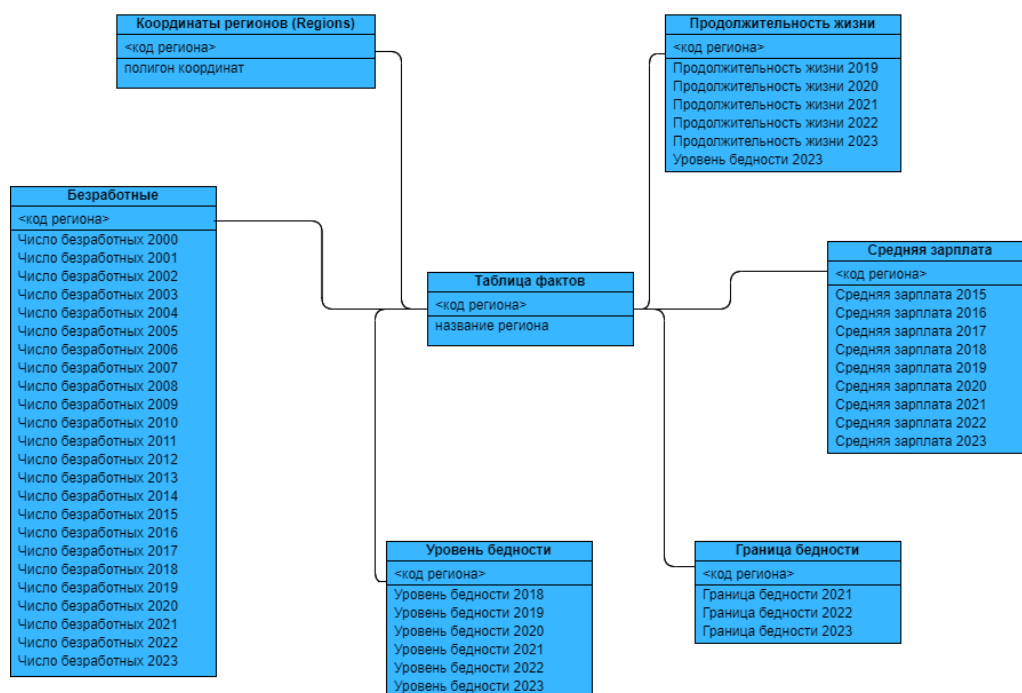


Рисунок 3.3. – Схема «Звездочка»

Для дальнейшей работы также нужно уточнить пару деталей. Во-первых, данные собирались в декабре 2024 года, поэтому в таблицах данные заканчиваются 2023 годом. Именно поэтому прогноз в следующих шагах будет на 2024 и 2025 годы. Во-вторых, перед загрузкой в PostgreSQL таблицы немного корректировались: удалились текстовые данные (пометки и названия Росстата), переименовывались столбцы.

В ходе работы в PostgreSQL была создана база данных под названием «risk_management» и загружены необходимые таблицы, заранее преобразованные в формат «csv». С помощью ERD можно посмотреть на созданную базу данных (рисунок 3.4).

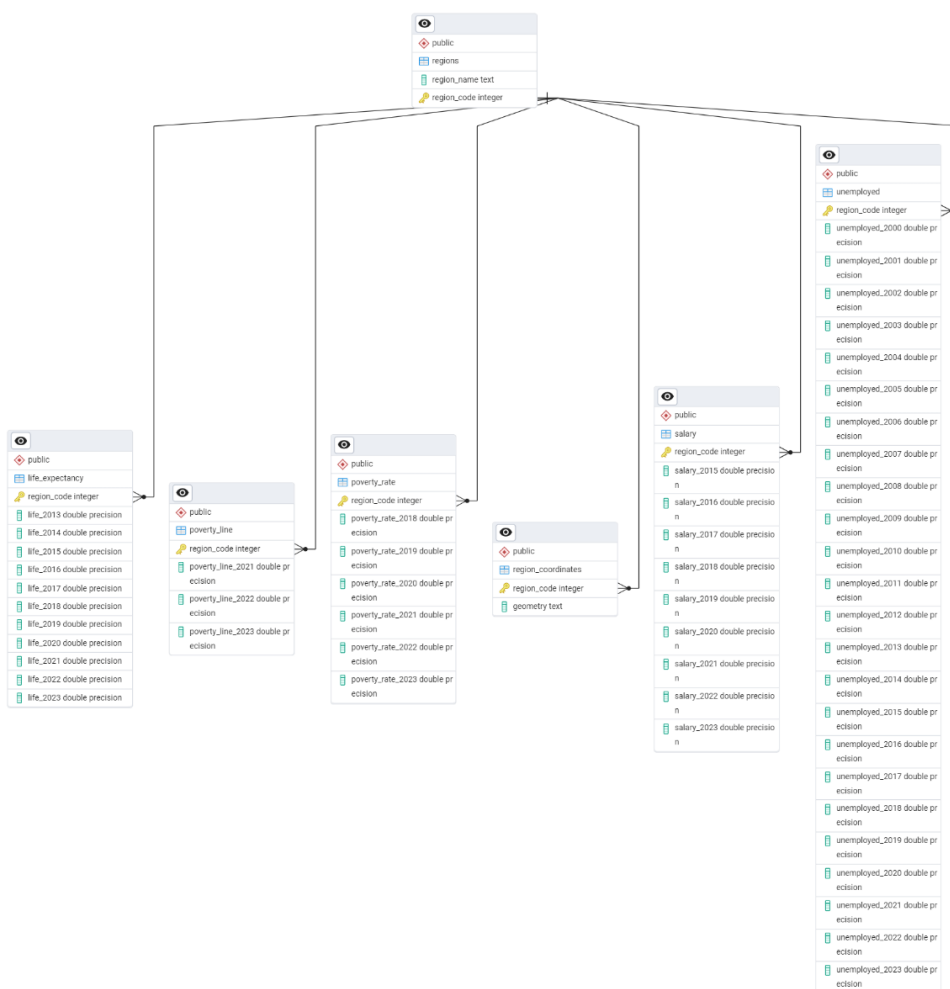


Рисунок 3.4. – ERD базы данных PostgreSQL «risk_management»

Пример того, как создавались «скелеты» в PostgreSQL для необходимых таблиц, можно увидеть на рисунке 3.5. Для этого использовался язык SQL с инструментом «Query Tool».

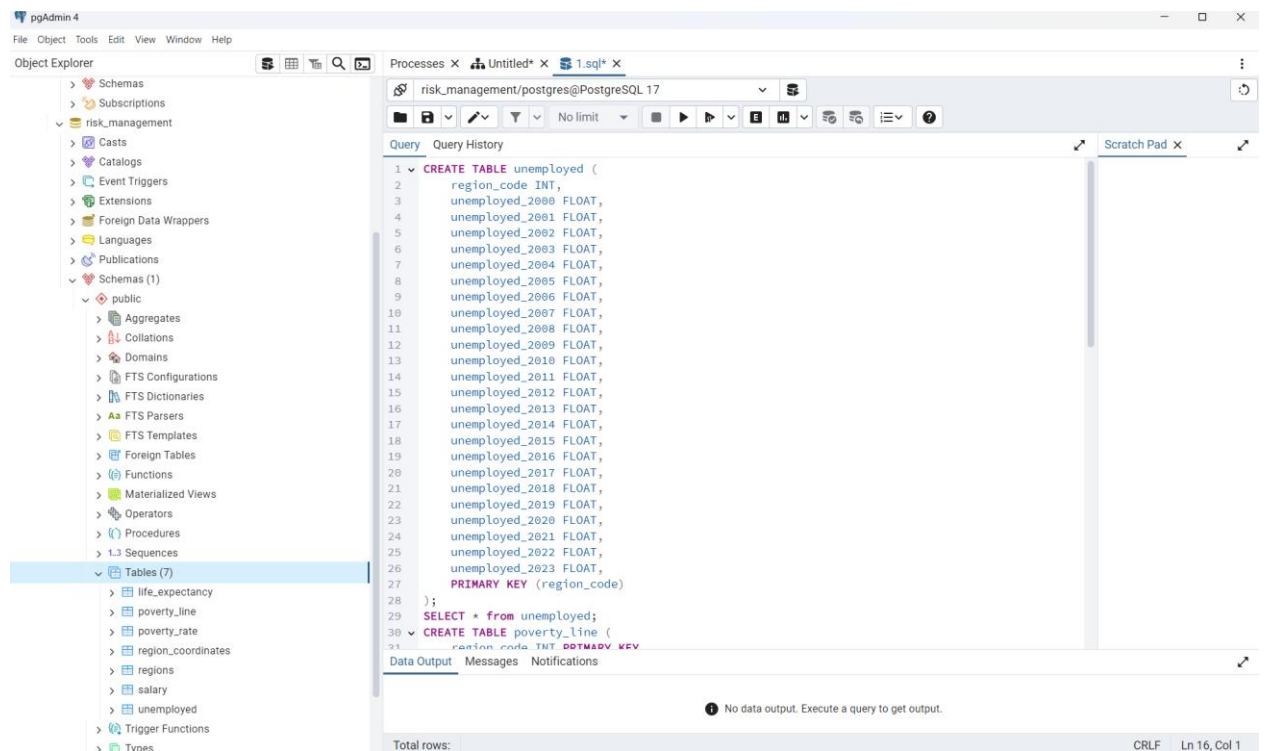


Рисунок 3.5. – Создание таблицы «Безработные»

Также можно посмотреть вывод одной из таблиц базы данных с помощью языка SQL (рисунок 3.6).

81 SELECT * from salary;

Data Output Messages Notifications

Showing rows: 1 to 85 Page No: 1 of 1

	region_code [PK] integer	salary_2015 double precision	salary_2016 double precision	salary_2017 double precision	salary_2018 double precision	salary_2019 double precision	salary_2020 double precision	salary_2021 double precision	salary_2022 double precision
1	31	23702	24444	25968	28434	30094	32296	34999	
2	32	19754	20092	21458	23606	24989	26077	28552	
3	33	21995	22807	24239	26781	27957	29769	32368	
4	36	22413	23396	24616	26820	28263	29723	32706	
5	37	18031	19335	20140	21943	22924	23965	25909	
6	40	27254	27949	30024	33329	34495	36567	39817	
7	44	19267	20208	21619	24389	26572	27431	30290	
8	46	22202	22628	23785	26087	27815	30075	33551	
9	48	23094	23369	24726	27285	29214	30734	33186	
10	50	38172	38830	41921	45377	46417	47677	51548	
11	57	19876	20968	22350	24632	26080	27701	30495	
12	62	23015	23797	25146	27800	29054	30927	33796	

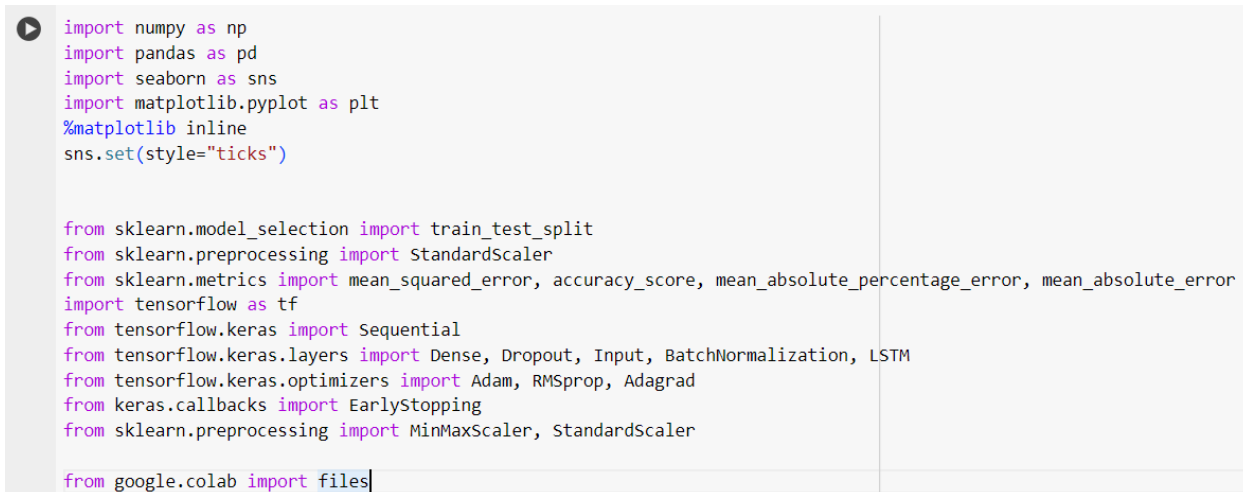
Total rows: 85 Query complete 00:00:00.291 CRLF Ln 81. Col 22

Рисунок 3.6. – Вывод таблицы «Средняя зарплата»

На данном этапе работу с PostgreSQL и созданием базы данных можно считать оконченной.

3.3. Анализ данных

После того, как была создана база данных с нужной информацией, можно начать анализ данных. В первую очередь необходимо подключить нужные библиотеки (рисунок 3.7).



```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, accuracy_score, mean_absolute_percentage_error, mean_absolute_error
import tensorflow as tf
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense, Dropout, Input, BatchNormalization, LSTM
from tensorflow.keras.optimizers import Adam, RMSprop, Adagrad
from keras.callbacks import EarlyStopping
from sklearn.preprocessing import MinMaxScaler, StandardScaler

from google.colab import files
```

Рисунок 3.7. – Загрузка необходимых библиотек Python

Библиотека NumPy позволяет работать с многомерными массивами и матрицами, она также содержит функции для выполнения математических операций. Библиотека Pandas предназначена для работы с данными. Она предоставляет структуры данных и функции для анализа и манипуляции данными. Библиотека Seaborn используется для визуализации. Модуль pyplot из библиотеки Matplotlib предоставляет функции для создания статических, анимационных и интерактивных визуализаций. `%matplotlib inline` позволяет визуализациям быть встроенными в вывод ячеек в Colab. Строка `«sns.set(style="ticks")»` устанавливает стиль визуализации Seaborn.

Класс `train_test_split` из библиотеки `scikit-learn` используется для разделения данных на обучающую и тестовую выборки. `StandardScaler` применяется для стандартизации признаков (приведение их к нулевому среднему и единичной дисперсии). Функции `mean_squared_error`, `accuracy_score`, `mean_absolute_percentage_error`, `mean_absolute_error` используются для оценки качества модели. Они позволят рассчитать среднеквадратичную ошибку, точность, среднюю абсолютную процентную ошибку и среднюю абсолютную ошибку.

Библиотека TensorFlow (`import tensorflow as tf`) является инструментом для построения и обучения нейронных сетей, а класс `Sequential` из Keras используется для создания последовательной нейронной сети, в которой слои идут друг за другом.

Классы `Dense`, `Dropout`, `Input`, `BatchNormalization`, `LSTM` из Keras:

- `Dense` – полносвязный слой, где каждый нейрон соединён со всеми нейронами предыдущего слоя;
- `Dropout` – это слой для регуляризации, отключающий случайные нейроны в процессе обучения;
- `Input` задаёт входные данные для сети;
- `BatchNormalization` – слой нормализации, который улучшает скорость и стабильность обучения;
- `LSTM` – слой для работы с временными рядами, который сохраняет информацию о последовательностях данных.

Классы `Adam`, `RMSprop`, `Adagrad` из Keras – это популярные оптимизаторы для настройки весов нейронной сети.

Класс `EarlyStopping`: используется для остановки обучения модели, если её производительность на валидационном наборе данных не улучшается в течение определённого количества эпох. `MinMaxScaler` и `StandardScaler` из `scikit-learn` используются для масштабирования данных, чтобы привести их значения в определённый диапазон (например, от 0 до 1) или стандартизировать их.

Библиотека `google.colab` (`from google.colab import files`) позволит скачать таблицы с новыми столбцами.

Далее была написана функция `forecast_with_metrics` (Приложение 1), которая прогнозирует данные (временные ряды) на два года вперед (2024 и 2025 года). Такой подход позволил сократить время и сам код для каждой из таблиц. Основное предназначение функции — создать прогнозы для показателей, используя метод нейронных сетей, а также оценить качество этих

прогнозов с помощью метрик, таких как среднеквадратичная ошибка (MSE) и точность (Accuracy).

В самом начале происходит поиск колонок, относящихся к годам, основываясь на заданном префиксе (`year_prefix`). Если таких колонок не найдено, программа выдает ошибку, так как прогнозировать без исходных данных невозможно. После этого создается копия исходной таблицы, чтобы сохранить результаты прогнозов без изменения оригинальных данных. Также создаются пустые списки для хранения значений MSE и Accuracy по каждому региону. Эти списки понадобятся для последующего вычисления средних значений. Для нормализации данных используется «MinMaxScale», который преобразует значения в диапазон от 0 до 1, чтобы ускорить и стабилизировать процесс обучения модели.

Для каждого региона данные временного ряда извлекаются и нормализуются. Затем формируется обучающая выборка: входные данные (X) представляют собой значения одного года, а выходные (y) – значения следующего. В случае отсутствия данных или невозможности сформировать обучающую выборку, цикл для текущего региона пропускается.

Модель для прогнозирования реализована на основе рекуррентной нейронной сети с использованием слоя LSTM. Этот слой особенно эффективен для анализа временных рядов, так как он способен учитывать предыдущие значения при прогнозировании. После добавления полносвязного слоя (Dense) модель компилируется с оптимизатором «adam» и функцией потерь «MSE». Она обучается на данных конкретного региона в течение 50 эпох.

После обучения модель прогнозирует значение на следующий год. Для этого используется последнее значение временного ряда текущего региона. Затем результат масштабируется обратно к исходному диапазону и добавляется в таблицу. Далее, используя предсказанное значение как новое «последнее», модель прогнозирует еще одно значение — на 2025 год, которое также добавляется в таблицу.

Качество прогноза оценивается по нескольким метрикам. MSE измеряет среднеквадратичную ошибку между реальными и предсказанными значениями, показывая, насколько точно модель предсказывает данные. Точность (Accuracy) рассчитывается как доля точек, где ошибка между реальными и предсказанными значениями меньше заданного порога (accuracy_threshold).

В конце функция выводит средние значения MSE и Accuracy для всех регионов, что позволяет оценить общую производительность модели. Итоговый результат возвращается в виде обновленной таблицы, где добавлены столбцы с прогнозами на 2024 и 2025 годы.

После этого подгружаются необходимые таблицы, добавляются таблицы с данными по алкоголизму и наркомании. Первым шагом проверяется наличие пустых значений («название_таблицы.isnull().sum()») – их в загружаемых таблицах нет. Далее функция обрабатывает таблицы со средней зарплатой, продолжительностью жизни, уровнем бедности, алкоголизмом и наркоманией. Прогноз для этих таблиц делается «изолированно», то есть без учета связей с другими. Данные по ошибкам представлены в таблице 3.1.

Таблица 3.1. – Показатели моделей

Таблица	Средняя MSE	Средняя Accuracy
Средняя зарплата	1124584129.85	0.00%
Продолжительность жизни	503.74	0.00%
Уровень бедности	133.61	0.00%
Алкоголизм	1605086.46	0.00%
Наркомания	57744.16	0.00%

Результаты с высокой средней MSE и нулевой точностью объясняются спецификой данных и ограничениями модели. Для средней зарплаты и алкоголизма высокие значения MSE связаны с масштабами показателей, где даже небольшие отклонения сильно влияют на ошибку. Продолжительность жизни и уровень бедности имеют более низкий MSE из-за ограниченного

диапазона значений, но точность остается нулевой из-за сложности выявления устойчивых закономерностей. К тому же, на данном этапе цель показать работу системы в целом, без привязки к точности предсказания.

Для наркомании MSE относительно низкий, но модель всё же не укладывается в заданный порог точности. Основные причины таких результатов – высокая вариативность данных, возможная шумность и отсутствие дополнительных факторов, которые могли бы улучшить прогноз. Это может указывать на дальнейшую необходимость доработки модели или расширения набора данных.

Далее происходит работа с таблицей безработных, чей прогноз будет осуществляться с учетом корреляции с другими данными. Для этого таблица объединяется с уже описанными в этой части работы данными, а также с таблицей с границей бедности. В ходе работы функции прогноз на два года вперед будет не только для числа безработных, но и для прожиточного минимума. До применения функции была выведена матрица корреляций (рисунок 3.9).

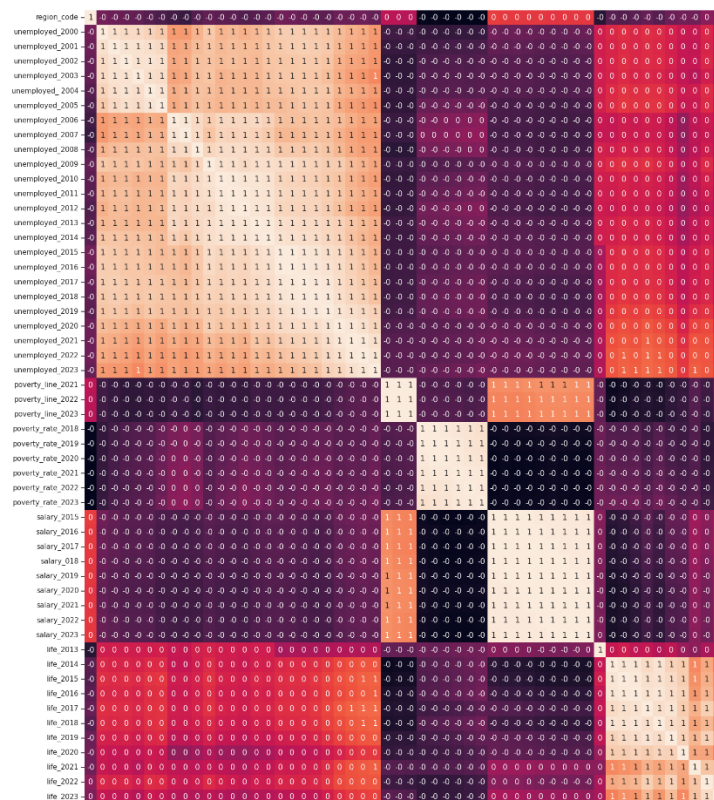


Рисунок 3.9. – Матрица корреляций общей таблицы

Из данной матрицы можно понять, что на численность безработный влияет граница бедности и средняя зарплата (чем они меньше, тем больше безработных), а также продолжительность жизни, однако они имеют слабую корреляцию. На границу бедности влияет численность безработных (отрицательная корреляция, чем больше безработных, тем ниже граница).

После применения функции `forecast_with_metrics` (ее было необходимо немного переписать для правильной работы с общей таблицей) (Приложение 2) результаты были таковыми:

- Средняя MSE: 1622290.72;
- Средняя Ассигура: 0.00%.

Далее произошла выгрузка таблиц с новыми столбцами с прогнозом на 2024 и 2025 годы. На этом этапе работа с Python была закончена.

3.4. Разработка VI-системы

Для начала работы с Yandex DataLens было создано подключение к новым таблицам под названием «РИСУР» (распределенная информационная система управления рисками), а также подключения «Алкоголизм» и «Наркомания».

Далее были образованы датасеты: Наркомания РФ (одна таблица), Заболеваемость алкоголизмом (одна таблица) и СУР (остальные таблицы). Рассмотрим подробнее последний датасет (рисунок 3.11).

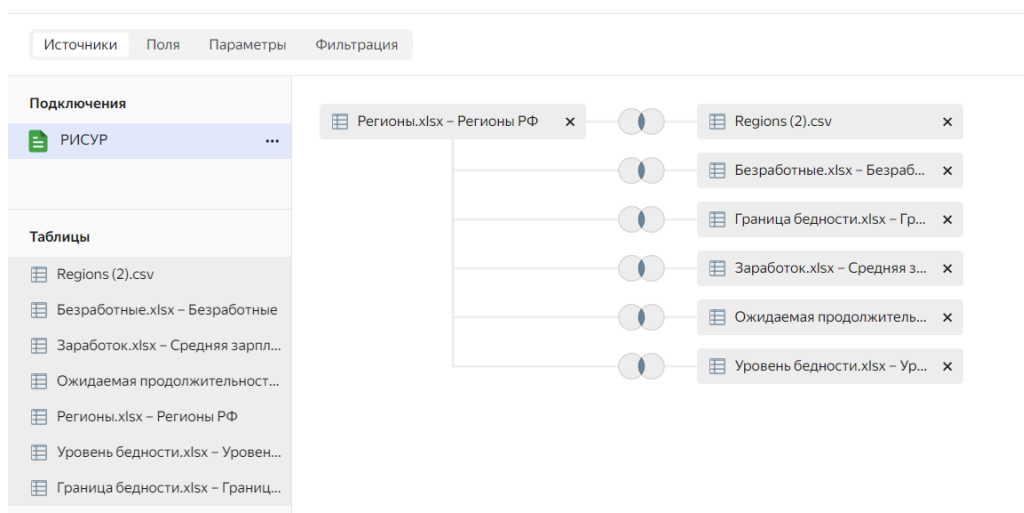


Рисунок 3.11. – Источники датасета «СУР»

Тут также была схема данных «Звездочка» и связь данных по коды субъекта РФ. После этого были созданы необходимые чарты – в основном это были карты (для них нужна была таблица координатами) (рисунок 3.12).

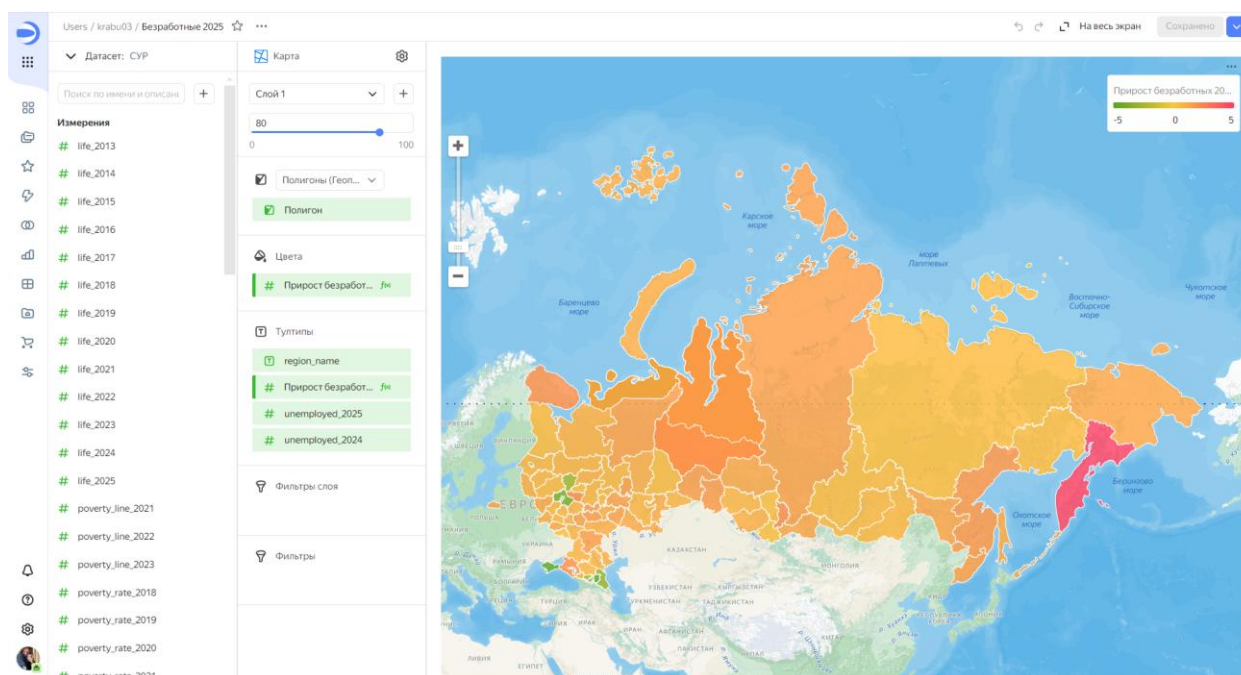


Рисунок 3.12. – Пример создания чарта по безработным 2025

Как можно заметить на рисунке 3.12, помимо использования тех полей, которые уже были в датасете, для каждого такого чарта использовалось написание функции «(рассматриваемый_год – предыдущий_год)/предыдущий_год * 100» (рисунок 3.13), для того чтобы смотреть на ситуацию в динамике.

Настройка поля

Прирост безработных 2025

Поле

1
$$\frac{[unemployed_2025] - [unemployed_2024]}{[unemployed_2024]} * 100$$

Рисунок 3.13. – Пример реализации функции динамики показателя

Всего было создано 18 чартов, 12 из которых в виде карты, остальные – в виде индикатора для показателей наркомании и алкоголизма, так как это показатели по все России.

В конечном итоге был создан итоговый дашборд «Система управления рисками» (Приложение 3), в котором информация сгруппирована по годам, т.е.

каждый показатель имеет два вида – информация на 2024 год и на 2025. На дашборде присутствуют все сделанные карты: 10 карт, которые сгруппированы по своей тематике, между ними можно переключаться. Также карты имеют подсказки как цветовые, так и тестовые – при наведении курсора на область, показывается:

- название субъекта РФ;
- показатель, по которому определяется риск (динамика);
- показатель этого года;
- показатель предыдущего года.

Также присутствуют 8 индикаторов для показателей наркомании и алкоголизма, которые также сгруппированы по признаку и году.

3.5. Оценка результатов

Интерпретируем дашборд, ставя цель выявить какой-либо риск на 2025 год. Посмотрим на карты с безработными: наибольший прирост безработных в 2024 году приходился на Тюменскую область, а наименьший – даже отрицательный – на Ямало-Ненецкий АО. В 2025 регионом риска, в котором ожидается самый большой прирост безработных в размере 7%, является Камчатский край.

Также Камчатский край в 2025 году имеет большой риск уменьшить свой средний заработок на 8,64%. Наименьшее уменьшение этого показателя наблюдается у Кировской области.

В 2024 году Чукотский АО мог увидеть увеличение уровня бедности больше остальных регионов, в 2025 году в зоне риска находятся Томская и Новосибирская области.

Ямало-Ненецкий АО в 2024 году мог наблюдать уменьшение прожиточного минимума. Самое большое уменьшение границы бедности в 2025 прогнозируется в Иркутской области.

Если говорить об ожидаемой продолжительности жизни, возможно в 2024 году, эти показателем Чукотский АО не мог похвастаться: он самый низкий – с 29% до 21. Примерно такая же ситуация и в 2025.

Говоря об общероссийской статистике: показатель алкоголизма растет быстрее, хотя и процент наркомании растет примерно с такой же скоростью.

В итоге, Камчатский край можно выделить как регион с наибольшими социально-экономическими рисками на 2025 год, включая прирост безработицы и снижение средней заработной платы. Также стоит обратить внимание на Томскую, Новосибирскую и Иркутскую области, которые демонстрируют негативные тенденции в других важных показателях.

Ссылка на материалы и на сам дашборд находится в Приложении 4.

Заключение

В рамках данного курсового проекта была реализована работа по концепции распределенной информационной системы построения модели угроз и управления гражданскими рисками. Платформа должна анализировать и прогнозировать ключевые социально-экономические показатели. Проект объединил в себе современные подходы к управлению данными, машинное обучение и визуализацию информации, что позволило создать инструмент, способный как выявить текущие риски, так и спрогнозировать их развитие в будущем.

На теоретическом уровне проект основывался на концепциях риск-менеджмента, анализа временных рядов и работы с социально-экономическими данными. Социально-экономические показатели, такие как уровень безработицы, бедности, средняя заработная плата, уровень алкоголизма и наркомании, имеют большое значение для оценки благосостояния регионов и всей страны. Их анализ позволяет выявить проблемы, требующие внимания со стороны государства и местных органов власти, а прогнозирование помогает принимать обоснованные управленческие решения.

Основная задача данной работы заключалась в описании концепции распределенной информационной системы управления рисками и построения модели угроз, а также в технической реализации для демонстрации работоспособности такой платформы.

В первой главе описывались теоретические задатки проекта – давались основные определения общего понимания системы управления рисками. Упоминались уже существующие решения в Российской Федерации, например на Федеральной таможенной службе. Также определялась правовая основа проекта.

Во-второй главе был сделан переход на описание непосредственной концепции РИСУР: вводились понятия, принцип работы системы и примеры

возможных угроз. Также перечислялись математические инструменты для реализации работы платформы, такие как алгоритмы классификации и кластеризации, байесовская математика и методы теории игр. Описывались возможные источники данных и планы реагирования на риски.

В третьей главе были описаны шаги по технической реализации демонстрации работы такой системы. Для хранения и обработки данных использовалась реляционная база данных PostgreSQL. Она основывалась на девяти таблицах, включающих информацию о безработице, уровне бедности, средних заработках, уровне алкоголизма и наркомании, продолжительности жизни, координатах регионов и других ключевых данных.

Прогнозирование осуществлялось с использованием нейронной сети на основе библиотеки TensorFlow. Реализованная функция прогнозирования учитывала исторические данные для каждого региона, нормализовала их и обучала модель для предсказания значений на 2024 и 2025 годы. Для создания дашборда был использован Yandex DataLens. Этот инструмент позволил построить интерактивные карты, графики и диаграммы, предоставляющие удобный доступ к информации о каждом регионе.

На основании прогнозов были выделены регионы с наибольшими рисками. Например, наибольший прирост безработицы в 2025 году ожидается в Камчатском крае (+7%), где также прогнозируется снижение средней зарплаты на 8,64%. В области бедности рисковыми регионами оказались Томская и Новосибирская области, а Иркутская область может столкнуться с самым большим снижением границы бедности.

Созданный инструмент может быть использован для стратегического планирования на уровне государственного управления. В долгосрочной перспективе его можно доработать для более точных прогнозов и дополнить новыми данными, включая климатические, демографические и экономические факторы. Проект продемонстрировал, что грамотное использование баз данных, инструментов визуализации и методов машинного обучения является важным шагом для их минимизации и улучшения качества жизни граждан.

Список использованных источников

1. Таможенный кодекс Евразийского экономического союза (ред. от 29.05.2019) (приложение N 1 к Договору о Таможенном кодексе Евразийского экономического союза)
2. О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года: указ Президента РФ от 07.05.2024 № 309
3. О Стратегии национальной безопасности Российской Федерации: указ Президента Российской Федерации от 02.07.2021 г. № 400
4. О единой межведомственной информационно-статистической системе: постановление Правительства РФ от 26.05.2010 г. N 367 (с изменениями и дополнениями)
5. Об утверждении стратегии и тактики применения системы управления рисками, порядка сбора и обработки информации, проведения анализа и оценки рисков, разработки и реализации мер по управлению рисками (за исключением рисков в области ветеринарии, санитарно-эпидемиологического надзора и обеспечения карантина растений): приказ ФТС России от 18 августа 2015 г. N 1677
6. Бюджет для граждан 2024–2026 к проекту федерального закона о федеральном бюджете на 2024 год и на плановый период 2025 и 2026 годов / Минфин России – 2023 – Режим доступа: https://minfin.gov.ru/common/upload/library/2023/10/main/0749_Budget_2024-2026_corr.pdf (дата обращения: 29.10.2024)
7. Advantages and Disadvantages Of Distributed Systems [Электронный ресурс] / Geeks for geeks – 06.08.2024 – Режим доступа: <https://www.geeksforgeeks.org/advantages-and-disadvantages-of-distributed-systems/> (дата обращения: 28.10.2024)
8. Гвоздева Е.А., Сорокин А.В. Риск-менеджмент: Учебное пособие для студентов всех форм обучения направления подготовки «Экономика».

- Издание 2-е дополненное и исправленное / Рубцовский индустриальный институт. – Рубцовск, 2021 – 84 с.
9. Егоров Александр Вадимович, Куприянова Наталия Игоревна Особенности методов кластеризации данных // Известия ЮФУ. Технические науки. 2011. №11. Режим доступа: <https://cyberleninka.ru/article/n/osobennosti-metodov-klasterizatsii-dannyh> (дата обращения: 31.10.2024)
 10. Жилов Руслан Альбердович Интеллектуальные методы кластеризации данных // Известия КБНЦ РАН. 2023. №6 (116). Режим доступа: <https://cyberleninka.ru/article/n/intellektualnye-metody-klasterizatsii-dannyh> (дата обращения: 31.10.2024)
 11. Кремлев, А.Г. К79 Основные понятия теории игр: учебное пособие / А.Г. Кремлев.— Екатеринбург : Изд-во Урал. ун-та, 2016.— 144 с
 12. Кузнецов В. Н. и др. Система принятия решения на основе теоремы Байеса //XXIII Санкт-петербургская международная конференция по интегрированным навигационным системам. – 2016. – С. 421-423.
 13. Ожегов С. И. Толковый словарь русского языка / С. И. Ожегов, Н. Ю. Шведова. – 4-е изд., доп. – Москва: Азбуковник, 2000 – 940 с.
 14. Утюганова В. В., Сердюк В. С., Фомин А. И. Прогнозирование и оценка профессиональных рисков в горной отрасли с применением теоремы Байеса //Безопасность труда в промышленности. – 2021. – №. 1. – С. 79-87.
 15. Шерстнев П. А., Липинский Л. В. Вычисление вектора документа с использованием меры TF-IDF // Актуальные проблемы авиации и космонавтики. 2021. №. Режим доступа: <https://cyberleninka.ru/article/n/vychislenie-vektora-dokumenta-s-ispolzovaniem-mery-tf-idf> (дата обращения: 31.10.2024)
 16. Банк документов Президента России [Электронный ресурс] / Кремль – 2024 – Режим доступа:

- <http://www.kremlin.ru/acts/bank/search?title=доктрина> (дата обращения: 27.10.2024)
17. Бахтиаров С. С. Теорема Байеса как метод определения атак типа отказ в обслуживании // Символ науки. 2018. №7. Режим доступа: <https://cyberleninka.ru/article/n/teorema-bayesa-kak-metod-opredeleniya-atak-tipa-otkaz-v-obsluzhivanii> (дата обращения: 31.10.2024)
18. Обзор практики применения исполнительными органами Ханты-Мансийского автономного округа – Югры системы оценки и управления рисками причинения вреда (ущерба) охраняемым законом ценностям при осуществлении регионального государственного контроля (надзора) в 2022 году [Электронный ресурс] / Департамент экономического развития Ханты-Мансийского автономного округа – Югры – Режим доступа: <https://depeconom.admhmao.ru/upload/iblock/62a/bs1elppt086w8761ui3ez7mgllhm5ajk9/Obzor-praktiki-primeneniya-ispolnitelnymi-organami-sistemy-upravleniya-ri....pdf> (дата обращения: 27.10.2024)
19. Павлова Оксана Сергеевна Исторический обзор становления риск-менеджмента как науки // Известия СПбГЭУ. 2010. №3. URL: <https://cyberleninka.ru/article/n/istoricheskiy-obzor-stanovleniya-risk-menedzhmenta-kak-nauki-1> (дата обращения: 02.12.2024)
20. Перечень государственных программ, национальных и федеральных проектов, приоритетных программ и проектов в Российской Федерации [Электронный ресурс] / КонсультантПлюс – Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_310251/ (дата обращения: 29.10.2024)
21. Платформа BI Yandex DataLens [Электронный ресурс] / Yandex – Режим доступа: <https://datalens.yandex.cloud/> (дата обращения: 10.01.2025)
22. Среда программирования Google Colab [Электронный ресурс] / Google – Режим доступа: <https://colab.research.google.com/> (дата обращения: 10.01.2025)

23. Старовойтов Владимир Гаврилович Внедрение системы управления рисками в федеральных органах власти (на примере Федеральной таможенной службы) [Электронный ресурс] / Финансовый университет – Режим доступа: http://www.fa.ru/org/science/eipieb/Documents/about/Старовойтов_%2020-02-2018.pdf (дата обращения: 24.10.2024)
24. СУБД PostgreSQL [Электронный ресурс] / PostgreSQL – Режим доступа: <https://www.postgresql.org/> (дата обращения: 10.01.2025)
25. Федеральная служба государственной статистики [Электронный ресурс] / Федеральная служба государственной статистики – Режим доступа: <https://rosstat.gov.ru/> (дата обращения: 09.01.2025)
26. Федеральная таможенная служба [Электронный ресурс] / Федеральная таможенная служба – 2024 – Режим доступа: https://customs.gov.ru/uchastnikam-ved/spravochnaya-informacziya/Obshchie_svedeniya_o_primenenii_tamozhennymi_organami_sistemy_upravleniya_riskami (дата обращения: 24.10.2024)
27. Цифровая экономика РФ [Электронный ресурс] / Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации – Режим доступа: <https://digital.gov.ru/ru/activity/directions/858/> (дата обращения: 27.10.2024)

```

def forecast_with_metrics(data, id_col, year_prefix, accuracy_threshold=0.1):
    year_cols = [col for col in data.columns if year_prefix in col]
    if not year_cols:
        raise ValueError(f"Столбцы с префиксом '{year_prefix}' не найдены.")
    updated_data = data.copy()
    mse_list = []
    accuracy_list = []
    scaler = MinMaxScaler()
    for idx, row in data.iterrows():
        region_data = row[year_cols].values.reshape(-1, 1)
        if len(region_data) == 0:
            continue
        scaled_data = scaler.fit_transform(region_data)
        X, y = [], []
        for i in range(len(scaled_data) - 1):
            X.append(scaled_data[i:i + 1])
            y.append(scaled_data[i + 1])
        X, y = np.array(X), np.array(y)
        if len(X) == 0:
            continue
        model = Sequential([
            LSTM(50, activation='relu', input_shape=(X.shape[1], X.shape[2])),
            Dense(1)
        ])
        model.compile(optimizer='adam', loss='mse')
        model.fit(X, y, epochs=50, verbose=0)
        last_value = scaled_data[-1].reshape(1, 1, 1)
        next_value_scaled = model.predict(last_value)[0][0]
        next_value = scaler.inverse_transform([[next_value_scaled]])[0][0]
        next_year_2024 = int(year_cols[-1].split('_')[-1]) + 1
        updated_data.loc[idx, f"{year_prefix}_{next_year_2024}"] = next_value
        next_value_scaled_2024 = scaler.transform([[next_value]])[0][0]
        next_value_scaled_2024 = np.array([[next_value_scaled_2024]])
        next_value_scaled_2025 = model.predict(next_value_scaled_2024)[0][0]
        next_value_2025 = scaler.inverse_transform([[next_value_scaled_2025]])[0][0]
        next_year_2025 = next_year_2024 + 1
        updated_data.loc[idx, f"{year_prefix}_{next_year_2025}"] = next_value_2025
        y_pred_scaled = model.predict(X).flatten()
        y_pred = scaler.inverse_transform(y_pred_scaled.reshape(-1, 1)).flatten()
        mse = mean_squared_error(y.flatten(), y_pred)
        mse_list.append(mse)
        accuracy = np.mean(np.abs(y.flatten() - y_pred) < accuracy_threshold)
        accuracy_list.append(accuracy)
    avg_mse = np.mean(mse_list)
    avg_accuracy = np.mean(accuracy_list)
    print(f"Средний MSE для {year_prefix}: {avg_mse:.2f}")
    print(f"Средняя Accuracy для {year_prefix}: {avg_accuracy * 100:.2f}%")
    return updated_data

```

Рисунок 3.8. – Функция для прогноза на два года вперед

```

def forecast_with_metrics_un(data, id_col, year_prefixes, accuracy_threshold=0.1):
    updated_data = data.copy()
    for year_prefix in year_prefixes:
        year_cols = [col for col in data.columns if year_prefix in col]
        if not year_cols:
            raise ValueError(f"Столбцы с префиксом '{year_prefix}' не найдены.")
        mse_list = []
        accuracy_list = []
        scaler = MinMaxScaler()
        for idx, row in data.iterrows():
            region_data = row[year_cols].values.reshape(-1, 1)
            if len(region_data) == 0:
                continue
            scaled_data = scaler.fit_transform(region_data)
            X, y = [], []
            for i in range(len(scaled_data) - 1):
                X.append(scaled_data[i:i+1])
                y.append(scaled_data[i+1])
            X, y = np.array(X), np.array(y)
            if len(X) == 0:
                continue
            model = Sequential([
                LSTM(50, activation='relu', input_shape=(X.shape[1], X.shape[2])),
                Dense(1)
            ])
            model.compile(optimizer='adam', loss='mse')
            model.fit(X, y, epochs=50, verbose=0)
            last_value = scaled_data[-1].reshape(1, 1, 1)
            next_value_scaled = model.predict(last_value)[0][0]
            next_value = scaler.inverse_transform([[next_value_scaled]])[0][0]
            next_year_2024 = int(year_cols[-1].split('_')[-1]) + 1
            updated_data.loc[idx, f"{year_prefix}_{next_year_2024}"] = next_value
            next_value_scaled_2024 = scaler.transform([[next_value]])[0][0]
            next_value_scaled_2024 = np.array([[next_value_scaled_2024]])
            next_value_scaled_2025 = model.predict(next_value_scaled_2024)[0][0]
            next_value_scaled_2025 = scaler.inverse_transform([[next_value_scaled_2025]])[0][0]
            next_year_2025 = next_year_2024 + 1
            updated_data.loc[idx, f"{year_prefix}_{next_year_2025}"] = next_value_scaled_2025
            y_pred_scaled = model.predict(X).flatten()
            y_pred = scaler.inverse_transform(y_pred_scaled.reshape(-1, 1)).flatten()
            mse = mean_squared_error(y.flatten(), y_pred)
            mse_list.append(mse)
            accuracy = np.mean(np.abs(y.flatten() - y_pred) < accuracy_threshold)
            accuracy_list.append(accuracy)
        avg_mse = np.mean(mse_list)
        avg_accuracy = np.mean(accuracy_list)
        print(f"Средний MSE для {year_prefix}: {avg_mse:.2f}")
        print(f"Средняя Accuracy для {year_prefix}: {avg_accuracy * 100:.2f}%")
    return updated_data

```

Рисунок 3.10. – Функция для прогноза на два года вперед для общей таблицы

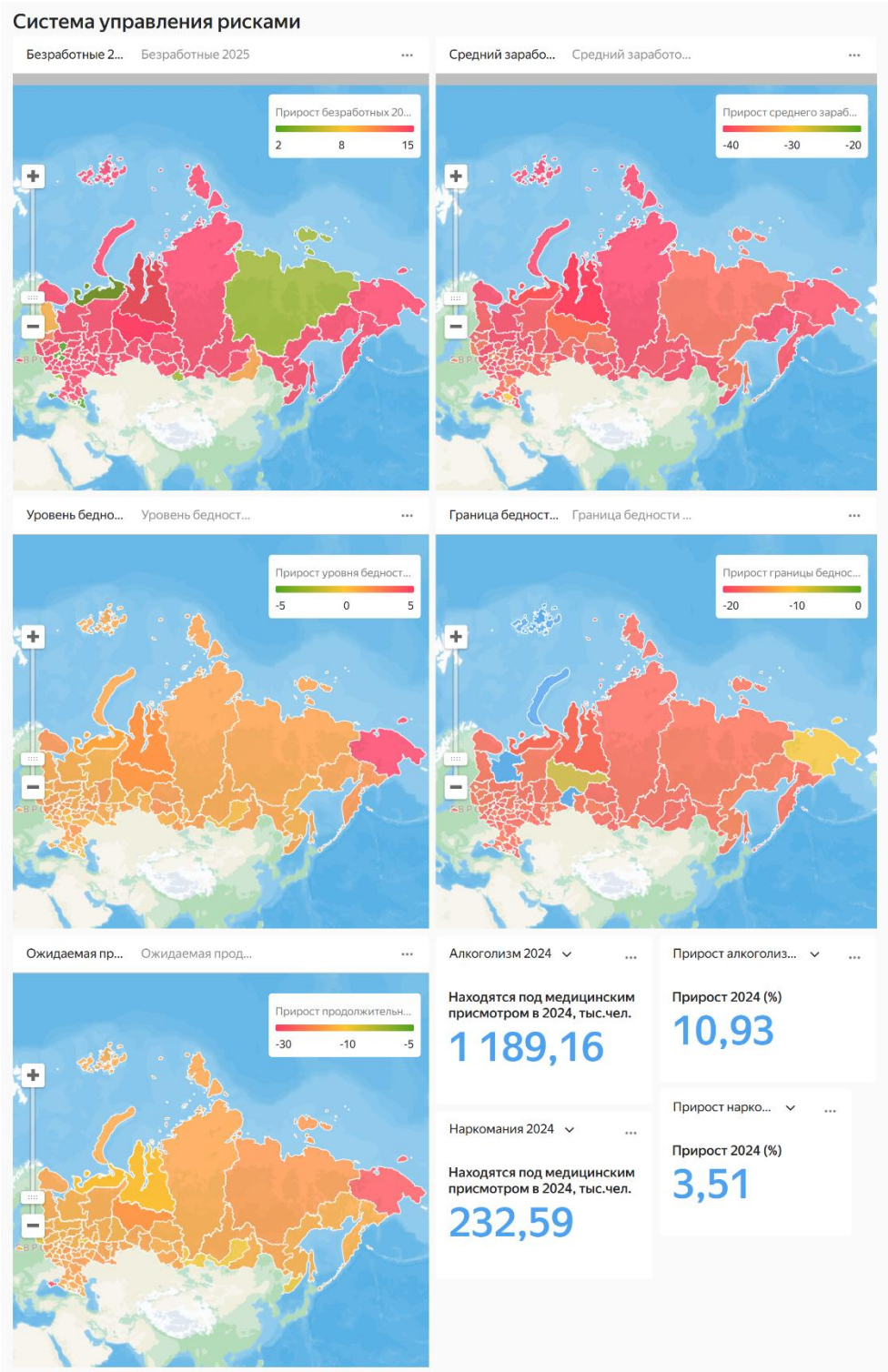


Рисунок 3.14. – Дашборд Система управления рисками

Приложение 4

Ссылка на Яндекс.Диск с материалами:

<https://disk.yandex.ru/d/Ya44w1RnVRujfg>