

ВОРОНЕЖСКИЙ ИНСТИТУТ ВЫСОКИХ ТЕХНОЛОГИЙ –
АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ
ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
(ВИВТ – АНОО ВО)

На правах рукописи



Ломаков Андрей Владимирович

**ИНТЕЛЛЕКТУАЛИЗАЦИЯ ПРОЦЕССОВ УПРАВЛЕНИЯ В РАМКАХ
РЕГИОНАЛЬНОЙ ОТРАСЛЕВОЙ ОРГАНИЗАЦИОННОЙ СИСТЕМЫ
НА ОСНОВЕ СРЕДСТВ ПРЕДИКТИВНОГО АНАЛИЗА
РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ**

Специальность: 2.3.4. Управление в организационных системах

ДИССЕРТАЦИЯ
на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор технических наук, доцент
Гусев Павел Юрьевич

Воронеж – 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
ГЛАВА 1. ПУТИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ РЕСУРСНЫМ ОБЕСПЕЧЕНИЕМ РЕГИОНАЛЬНОЙ ОТРАСЛЕВОЙ ОРГАНИЗАЦИОННОЙ СИСТЕМЫ НА ОСНОВЕ РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ.....	
1.1. Особенности использования ретроспективной информации при управлении в региональной отраслевой организационной системе.....	15
1.2. Повышение эффективности управления за счет интеграции результатов предиктивного анализа и оптимизации	23
1.3. Структурное моделирование региональной отраслевой организационной системы и интеллектуальной поддержки принятия решений при управлении ею	40
Выводы первой главы.....	45
ГЛАВА 2. ФОРМИРОВАНИЕ РЕЗУЛЬТАТОВ ПРЕДИКТИВНОГО АНАЛИЗА РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ НА ОСНОВЕ ОБЪЕДИНЕНИЯ ВИЗУАЛЬНОГО МОДЕЛИРОВАНИЯ И МОДЕЛИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ	
2.1. Структура объединенной алгоритмической схемы визуального моделирования и моделирования с использованием машинного обучения.....	48
2.2. Формирование ансамбля прогностических моделей показателей эффективности и объемов дополнительного ресурсного обучения	53
2.3. Прогностическое моделирование интегрального эффекта от использования дополнительного ресурсного обеспечения	69
Выводы второй главы	71
ГЛАВА 3. РАЗРАБОТКА ПРОЦЕДУР ИНТЕГРАЦИИ РЕЗУЛЬТАТОВ ПРЕДИКТИВНОГО АНАЛИЗА РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ В ПРОЦЕСС ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ НА ОСНОВЕ ОПТИМИЗАЦИИ.....	
3.1. Управление распределением ресурсного обеспечения между группами населения региональной отраслевой организационной системы.....	74
3.2. Управление распределением ресурсного обеспечения между территориальными образованиями региональной отраслевой организационной системы	83
3.3. Управление распределением ресурсного обеспечения между временными периодами в рамках горизонта прогнозирования	87
Выводы третьей главы.....	90

ГЛАВА 4. ПРИМЕНЕНИЕ РАЗРАБОТАННЫХ СРЕДСТВ ИНТЕЛЛЕКТУАЛИЗАЦИИ ПРИНЯТИЯ РЕШЕНИЙ В ПРАКТИКЕ УПРАВЛЕНИЯ РЕГИОНАЛЬНОЙ СИСТЕМОЙ ЗДРАВООХРАНЕНИЯ.....	92
4.1. Структура программного обеспечения.....	93
4.2. Визуальное и прогностическое моделирование заболеваемости артериальной гипертензией лиц старших возрастных групп и их диспансеризации в системе здравоохранения Воронежской области.....	97
4.3. Оценка эффективности применения разработанных средств при управлении процессом диспансеризации на основе вычислительного эксперимента.....	107
Выводы четвертой главы.....	116
ЗАКЛЮЧЕНИЕ	118
СПИСОК ЛИТЕРАТУРЫ	121
ПРИЛОЖЕНИЕ А. Акты внедрения.....	131
ПРИЛОЖЕНИЕ Б. Свидетельства о регистрации программ для ЭВМ	136
ПРИЛОЖЕНИЕ В. Сравнение результатов работы методов прогнозирования	138

ВВЕДЕНИЕ

Актуальность темы. В деятельности отраслевых региональных организационных систем (здравоохранение, образование, социальное обеспечение) важное место занимает ежегодный статистический учет, выполняемый по определенным формам и требованиям. В результате этой деятельности накапливается большой объем ретроспективной информации. В практике управления редко используют длительные наблюдения, ориентируясь на использование ретроспективных данных, ближайших по времени к плановому периоду. В теоретических исследованиях, связанных с повышением эффективности управления в региональных отраслевых организационных системах, рассматривается возможность принятия управленческих решений на основе моделирования, экспертного оценивания и оптимизации, но при этом повторяется тот же принцип использования временной информации. Указанные исследования отражены в трудах отечественных ученых С.А. Баркалова, А.М. Бершадского, В.Н. Буркова, М.Б. Гузаирова, В.А. Зернова, Я.Е. Львовича, В.А. Новикова, А.Н. Райкова, Ю.С. Сахарова, М.В. Щербакова и др.

Однако в них не делается акцент на возможности, которые предоставляют современные методы предиктивного анализа накопленных данных применительно к ретроспективной информации. Одним из аспектов управления в региональных отраслевых организационных системах, для которого целесообразно использовать результаты предиктивного анализа, является управление распределением ресурсного обеспечения. При этом реализация управленческих действий связана с рядом особенностей:

- принятие решения по трем составляющим распределения ресурса: группам населения региона, территориальным образованиям, временным периодам в рамках горизонта планирования;

- требования сбалансированности решений при распределении ресурсного обеспечения по трем составляющим с плановым объемом, заданным управляющим центром;
- учет множества показателей, фиксируемых региональной статистикой и характеризующих, с одной стороны, объемы использования ресурсного обеспечения, а с другой – показатели эффективности деятельности управляющего центра;
- преимущественная ориентация на экспертные оценки прошлого опыта управления ресурсным обеспечением при предсказании результатов функционирования системы в будущие временные периоды;
- достаточность данных статистического учета для перехода от оценок по множеству показателей эффективности к оценке интегрального эффекта от использования ресурсного обеспечения.

В условиях, связанных с перечисленными особенностями, традиционный экспертный подход не позволяет выбрать наиболее эффективный вариант управленческого решения. Поддержка экспертных решений процесса управления путем последовательного применения предиктивного анализа ретроспективной информации и методов интеллектуализации полностью не обеспечивает достижение лучшего результата.

Таким образом, актуальность темы определяется необходимостью интеграции средств предиктивного анализа ретроспективной информации в процесс управления распределением ресурсного обеспечения в региональной отраслевой организационной системе на основе средств интеллектуализации принятия решений.

Работа выполнена в рамках основного научного направления Воронежского института высоких технологий – автономной некоммерческой образовательной организации высшего образования: «Фундаментальные и прикладные исследования по разработке и совершенствованию информационных технологий, моделей, методов и средств автоматизации и управления техническими,

технологическими, экономическими и социальными процессами и производствами».

Целью работы является повышение эффективности управления в региональной отраслевой организационной системе на основе интеграции результатов предиктивного анализа ретроспективной информации в процесс принятия управленческих решений с использованием интеллектуальных методов.

Для достижения поставленной цели необходимо решить следующие задачи:

- рассмотреть пути повышения эффективности управления ресурсным обеспечением в региональной отраслевой организационной системе с использованием ретроспективной информации на основе двухуровневой схемы принятия решений и интеллектуализации процесса управления;
- разработать алгоритм предиктивного анализа, объединяющий визуальное моделирование и моделирование на основе машинного обучения с выбором методов, наиболее эффективных для обработки характерных выборок ретроспективной информации;
- разработать процедуры интеграции результатов предиктивного анализа при интеллектуализации управления распределением ресурсного обеспечения в региональной отраслевой организационной системе и их реализации в процесс принятия решений;
- разработать программный комплекс поддержки принятия управленческих решений на основе предиктивного анализа ретроспективной информации;
- оценить эффективность применения разработанных средств на примере управления процессом диспансеризации в региональной отраслевой организационной системе здравоохранения.

Объект исследования – процесс управления в региональной отраслевой организационной системе с использованием ретроспективной информации.

Предмет исследования – эффективность применения предиктивного анализа ретроспективной информации для оптимизации управления в региональной отраслевой организационной системе.

Методы исследования. Для решения поставленных задач использовались основные положения теории управления, методы предиктивной аналитики, методы классификации, модели и методы многоальтернативной оптимизации, экспертного оценивания.

Соответствие диссертации паспорту специальности. Содержание диссертации соответствует следующим пунктам паспорта научной специальности 2.3.4 «Управление в организационных системах»:

- п. 5 «Разработка методов получения данных и идентификации моделей, прогнозирования и управления организационными системами на основе ретроспективной, текущей и экспертной информации»;
- п. 8 «Разработка проблемно-ориентированных систем управления и оптимизации организационных систем»;
- п. 9 «Разработка методов и алгоритмов интеллектуальной поддержки принятия управленческих решений в организационных системах».

Научная новизна. В диссертационной работе получены следующие результаты, характеризующиеся научной новизной:

1. Структурная модель процессов управления в региональной отраслевой организационной системе, отличающаяся компонентами и связями между ними, двухуровневой схемой принятия решений на основе интеллектуальных методов и обеспечивающая сбалансированное распределение ресурсного обеспечения по трем составляющим в соответствии с требованиями управляющего центра.

2. Алгоритм предиктивного анализа ретроспективной информации, отличающийся объединением визуального моделирования и моделирования с использованием машинного обучения при прогнозировании функционирования региональной отраслевой организационной системы на будущие временные периоды с учетом вариативного использования данных по показателям эффективности и обеспечивающий формирование результатов для интеграции в процесс принятия управленческих решений.

3. Оптимационные модели интеллектуализации процесса поддержки принятия управленческих решений, отличающиеся преобразованиями прогностических оценок в параметры зависимостей экстремальных требований от оптимизируемых переменных и обеспечивающие принятие управленческих решений на основе многоальтернативной оптимизации.

4. Процедура принятия окончательного решения по распределению ресурсного обеспечения между временными периодами в рамках горизонта планирования, отличающаяся двухуровневым перебором при максимизации прогностических оценок функции интегрального эффекта и обеспечивающая коррекцию управленческих решений, полученных на основе многоальтернативной оптимизации.

5. Структура программного комплекса, отличающаяся реализацией механизма встраивания в отраслевые организационные системы и обеспечивающая преобразование и использование ретроспективной информации при принятии управленческих решений на основе результатов предиктивного анализа.

Теоретическая значимость заключается в развитии методов оптимизации при управлении в региональной отраслевой организационной системе путем интеграции в процесс принятия решений результатов предиктивного анализа ретроспективной информации.

Практическая значимость заключается в:

- использовании разработанных моделей и процедур для интеллектуальной поддержки административных управленческих решений при управлении ресурсами отраслевых региональных организационных систем;
- обеспечении согласованности распределения планового ресурсного обеспечения между группами населения региона, территориальными образованиями и временными периодами в рамках горизонта перспективного планирования;

– применении разработанных программных средств совместно со стандартными программными продуктами при интеграции результатов предиктивного анализа в процесс принятия управленческих решений.

На разработанные программные средства получено 2 свидетельства о государственной регистрации в реестре Федеральной службы по интеллектуальной собственности (Роспатент).

Достоверность и обоснованность результатов подтверждается корректным использованием методов искусственного интеллекта и интеллектуальной оптимизации для формализованной постановки и решения задач предиктивного анализа, принятия решений при управлении ресурсным обеспечением в региональных отраслевых организационных системах, а также сравнительным анализом результатов моделирования и реальных данных.

Положения, выносимые на защиту

1. Структурная модель процессов управления в региональной отраслевой организационной системе позволяет установить компоненты и связи между ними при использовании двухуровневой схемы принятия решений на основе интеллектуальных методов распределения ресурсного обеспечения.

2. Алгоритм предиктивного анализа ретроспективной информации позволяет объединить визуальное моделирование, моделирование на основе машинного обучения и выбор наиболее точного метода прогнозирования функционирования региональной отраслевой организационной системы на будущие временные периоды с использованием тестовой выборки из массива ретроспективных данных.

Внедрение результатов работы

Результаты диссертационной работы внедрены в: 1) деятельность ООО «Актив Компьютерс» (г. Москва) в целях обработки потоков данных аудита для мониторинга аппаратных ресурсов на основе программного обеспечения классификации данных; 2) учебный процесс ВИВТ-АНОО ВО (г. Воронеж) по направлениям подготовки 09.03.01 Информатика и вычислительная техника (профиль подготовки «Автоматизированные системы обработки информации и управления») и 09.04.01 Информатика и вычислительная техника (программа

«Киберфизические системы и искусственный интеллект» профиль подготовки «Облачная и сетевая инфраструктура систем искусственного интеллекта»); 3) деятельность ФГБУЗ «Медико-санитарная часть №97» Федерального медико-биологического агентства РФ (г. Воронеж) при прогнозировании объемов диспансеризации населения.

Апробация работы. Результаты работы докладывались на следующих научных мероприятиях: III, IV Международном научно-техническом форуме «Современные технологии в науке и образовании» СТНО-2020, СТНО-2021 (г. Рязань, РГРТУ, 2020, 2021); VI научно-технической конференции магистрантов Рязанского государственного радиотехнического университета (г. Рязань, РГРТУ, 2020); XXV Юбилейной всероссийской научно-технической конференции студентов, молодых ученых и специалистов «Новые информационные технологии в научных исследованиях» НИТ-2020 (г. Рязань, РГРТУ, 2020); Международной научно-практической конференции «Интеллектуальные информационные системы» (г. Воронеж, ВГТУ, 2022, 2023, 2025); Международной научно-практической конференции «Актуальные проблемы инновационных систем информатизации и безопасности» (г. Воронеж, ВИВТ, 2023); XV межвузовской научно-практической конференции: «Проблемные вопросы и организация научно-исследовательской деятельности студентов» (г. Воронеж, ВИВТ, 2023); VIII Международной выставке изобретений и инноваций имени Николая Гавриловича Славянова (г. Воронеж, ВГУИТ, 2023); Международной научно-практической конференции студентов, аспирантов и молодых ученых «Молодежь в науке: экономика, технологии и инновации» (г. Воронеж, ВИВТ, 2023); Международной молодежной научной школе «Оптимизация и моделирование в автоматизированных системах» (г. Воронеж, ВГТУ, 2024, 2025), а также на научно-технических семинарах кафедры искусственного интеллекта и цифровых технологий ВГТУ и научного клуба ВИВТ с 2022 по 2025 гг.

Публикации. По материалам диссертационной работы опубликовано 17 научных работ, в том числе 4 – в научных журналах, рекомендованных ВАК РФ. Получено 2 свидетельства о регистрации программ для ЭВМ.

Личный вклад автора. В работах, опубликованных в соавторстве и приведенных в конце автореферата, личный вклад соискателя состоит: в исследовании, подборе и применении ряда методов для формирования результатов анализа медико-статистической информации на основе прогностического моделирования; в процедуре интеграции результатов предиктивной аналитики в процесс принятия управленческих решений; в подготовке наборов данных, проведении регрессионного моделирования, исследование корреляции и анализ полученных результатов; в проведении исследования областей применения и осуществлении разработки программного обеспечения; в исследовании возможностей применения технологий машинного обучения в учреждениях здравоохранения; в исследовании и формализации требований к целесообразности автоматизации на производственных предприятиях; в исследовании и подборе методов машинного обучения в целях повышения эффективности результатов предиктивного анализа; в формировании и подготовке обучающего набора данных на основе статистических данных, проведении обучения классификационных моделей; в разработке программного обеспечения и алгоритмов взаимодействия программных модулей настройки параметров классификаторов с основным модулем.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы из 92 наименований и трех приложений. Основная часть работы изложена на 120 страницах, содержит 32 рисунка, 14 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе проанализированы пути повышения эффективности управления ресурсным обеспечением региональной отраслевой организационной системы на основе ретроспективной информации. Введены обозначения, необходимые для формирования структурной модели, характеризующей связи управляющего центра с распределением ресурсного обеспечения и показателями эффективности управления. Рассмотрены возможности предиктивного анализа

ретроспективной информации для повышения эффективности управления в региональной отраслевой организационной системе. Отмечена особенность интеллектуальной оптимизации, заключающаяся в синтезировании зависимостей экстремальных и граничных требований от оптимизируемых переменных на основе машинного обучения. Сформулирована структурная модель интеллектуализации управления распределением ресурсного обеспечения в региональной отраслевой организационной системе на основе ретроспективной информации.

Вторая глава посвящена формированию результатов предиктивного анализа ретроспективной информации на основе объединения визуального моделирования и моделирования с использованием статистических методов и алгоритмов машинного обучения.

Предложена структура алгоритмической схемы объединения указанных видов моделирования и охарактеризованы компоненты этой схемы.

Показано, что визуальное моделирование основано на использовании эффективных механизмов наглядно-образной интуиции человека при сравнении графических представлений временных рядов.

С целью формирования описания прогностических моделей как результата предиктивного анализа выбраны полученные варианты учета выявленных тенденций при прогнозировании временных рядов с использованием трех групп видов машинного моделирования:

- функциональное моделирование;
- моделирование с использованием деревьев решений и случайного леса;
- нейросетевое моделирование с использованием многослойного перцептрона.

При анализе исследуемых методов прогностического моделирования были учтены уровни заболеваемости среди районов Воронежской области. Для разделения районов области на уровни была решена задача многоклассовой классификации.

Третья глава посвящена разработке процедур интеграции результатов предиктивного анализа ретроспективной информации в процесс принятия управлеченческих решений на основе средств интеллектуализации принятия решений.

В четвертой главе проведен анализ применения разработанных средств интеллектуализации процесса принятия управлеченческих решений в практике управления распределением ресурсного обеспечения в региональной отраслевой системе здравоохранения. Обоснована необходимость разработки проблемно-ориентированных программных средств и дана их характеристизация. Приведена структурная схема взаимодействия этих продуктов с библиотеками стандартных программ машинного обучения и оптимизации.

ГЛАВА 1. ПУТИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ РЕСУРСНЫМ ОБЕСПЕЧЕНИЕМ РЕГИОНАЛЬНОЙ ОТРАСЛЕВОЙ ОРГАНИЗАЦИОННОЙ СИСТЕМЫ НА ОСНОВЕ РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ

Возможности повышения эффективности управления распределением ресурсного обеспечения в региональных отраслевых организационных системах определяются особенностями принятия решений управляющим центром с ориентацией на данные ежегодного статистического учета. Эти особенности проявляются, прежде всего, в отражении в ретроспективных данных тенденций изменения показателей эффективности функционирования системы и объемом ресурсного обеспечения и их детализации по группам населения и территориальным образованиям региона.

Одним из путей повышения эффективности с учетом перечисленных особенностей является систематизация накопленных данных в виде временных рядов с последующим применением предиктивного анализа для прогноза их изменений и функциональных зависимостей между ними. Интеллектуализация процесса принятия решений достигается путем интеграции результатов предиктивного анализа в оптимизационные модели, отражающие экстремальные и граничные требования, которые определяют распределение ресурсного обеспечения в региональной организационной системе за счет достижения целей управляющего центра.

Формализованные методы интеллектуализации процесса принятия управленческих решений выявляются на этапе структурного моделирования, предшествующего предиктивному анализу и оптимизации. Для этого, во-первых, формируется структурная модель региональной отраслевой организационной системы, ориентированная на использование ретроспективной информации при

взаимодействии управляющего центра с компонентами системы; во-вторых, формируется структурная модель интеллектуальной поддержки принятия решений в рамках системы управления распределением ресурсного обеспечения в региональной отраслевой организационной системе.

1.1. Особенности использования ретроспективной информации при управлении в региональной отраслевой организационной системе

Под региональной отраслевой организационной системой понимается совокупность компонентов, связанных с принадлежностью к определенным группам населения и территориям проживания в регионе, которые объединены в организационное целое посредством взаимодействия с управляющим центром для достижения заданных характеристик жизнедеятельности.

Для оценки достижения заданных характеристик жизнедеятельности в отраслях региональной отраслевой организационной системы (здравоохранение, образование, социальная защита, промышленность, аграрный сектор) организован ежегодный статистический учет, который характеризуется как система сбора и обобщения информации о массовых явлениях, фактах и процессах с целью контроля и управления ими [1]. При организации статистического учета в практику введены показатели и классификации, которые применяются в анализе процессов, происходящих в отраслевых организационных системах, путем перехода к системе национальных счетов, используемой в целях отражения информации о социальных и экономических процессах в определенный период [2].

Определяющим для реализации единой государственной политики в сфере официального статистического учета является Федеральный закон от 29 ноября 2007 г. № 282-ФЗ «Об официальном статистическом учете и системе

государственной статистики в Российской Федерации», который регламентирует организацию статистических наблюдений и формирование системы показателей статистики процессов, происходящих в организационных системах. Результаты статистического наблюдения представляются в форме статистических показателей. Под показателем подразумевается обобщенная количественная характеристика качественно определенных социально-экономических явлений и процессов применительно к конкретным условиям территорий проживания и времени [3].

В целях формирования и использования интегрированных информационных ресурсов, собираемых в Федеральной службе государственной статистики, разработана информационно-справочная система метаинформации – каталог статистических показателей, применяемых в статистической практике. В него включены показатели, представленные в формах федерального государственного наблюдения, а также расчетные показатели, являющиеся результатом обработки первичных данных.

Выделяют следующие три основных блока, в которых сгруппированы показатели каталога:

1. Блок показателей, характеризующих факторы развития отраслевых организационных систем, в котором отражены источники социально-экономического развития (инвестиции, цены и тарифы, организационно-технический потенциал).

2. Блок показателей, характеризующих результаты развития отраслевых организационных систем, в котором отражены социально-экономические показатели (производство товаров, работ, услуг; степень заболеваемости населения; характеристики доходов работников и условий труда).

3. Блок показателей, характеризующих статистическую оценку степени реализации целей развития отраслевых организационных систем, в котором показатели отражены в качестве отношения фактических показателей к прогнозным значениям (интегральный эффект ресурсного обеспечения отраслевых организаций).

В мировой и российской статистической практике выделяют следующие критерии качества статистических данных [4]:

- востребованность – соответствие предоставляемой статистической информации целям и задачам статистического исследования с учетом стоимости и затраты времени на формирование данных;
- своевременность – соблюдение приемлемой продолжительности периода времени, в течение которого статистическая информация остается актуальной и востребованной;
- достоверность – величина отклонения истинных значений показателей статистики от прогнозных;
- доступность – возможность получения результатов статистического наблюдения широким кругом лиц;
- интерпретируемость – возможность соотнесения статистической информации с объективной реальностью;
- сопоставимость – возможность проведения сравнения статистической информации;
- согласованность – возможность объединения статистической информации различными способами для различных целей с высоким уровнем надежности.

Качество статистической информации достигается за счет того, что статистическая деятельность базируется на научно обоснованной методологии сбора и обработки данных статистики [5].

При этом в настоящее время наблюдается изменение подходов к решению проблемы качества статистической информации, которая ранее опиралась на общепризнанную точку зрения, что проблема некачественных данных должна решаться на уровне источника предоставляемых статистических данных, например, путем использования средств контроля при регистрации данных, в сторону того, что качество данных должно обеспечиваться системами обработки данных при их использовании [6]. В трудах отечественных ученых М.Г. Короткова

[7], М.Р. Терованесова [8], А.Ю. Лабинского [9], Е.М. Богдановой [10] отражены подходы к обработке имеющихся статистических данных.

Рассмотрим статистический учет в одной из характерных региональных отраслевых организационных систем – здравоохранении. В региональных организационных системах здравоохранения существуют единые формы отчетности статистической информации, а также единые формы учета и инструкции по их заполнению. Они позволяют обобщать и сравнивать результаты медико-статистических данных здравоохранения по субъектам, городам и районам.

К принципиальным особенностям статистического учета в здравоохранении относятся:

- связь региональных организационных систем здравоохранения с управляющим центром;
- применение данных статистики в целях формирования планов мероприятий и оценки интегрального эффекта ресурсного обеспечения;
- взаимосвязь с практическими задачами региональной отраслевой организационной системы здравоохранения в целях выявления путей дальнейшего развития.

Достоверный медико-статистический анализ напрямую зависит от качества сбора первичных статистических данных в региональных организационных системах здравоохранения: их достоверности, актуальности, информативности, полноты [11]. Ключевым инструментом получения статистических данных в целях принятия правильных управленческих решений в учреждениях здравоохранения является система официальной статистической отчетности. При этом в отечественной системе государственной отчетности в сфере здравоохранения существует ряд проблем, связанных с отсутствием необходимых данных, в частности, о результатах экономической деятельности региональных организационных систем здравоохранения [12], в связи с чем отмечается необходимость приведения медико-статистической информации в соответствие

практическим задачам [13-14]. Выделяют следующие недостатки существующей статистической отчетности [15]:

- дублирование и пересечение показателей в формах в части средств обязательного медицинского страхования;
- неудобный вид данных и набор показателей для предоставления и анализа информации;
- отсутствие взаимосвязи между показателями статистической отчетности и внутриорганизационным управленческим учетом.

Данные недостатки предлагаются разрешить выработкой подходов по уменьшению нагрузки на региональные отраслевые организационные системы здравоохранения путем создания обособленных организаций анализа медико-статистической информации [16], а также аутсорсинга формирования баз данных в здравоохранении [17].

Определяющую функцию в осуществлении процесса сбора, обработки и хранения статистических данных о заболеваемости населения и функционирования региональной организационной системы здравоохранения выполняет первичное звено медико-статистической службы – кабинеты отделения статистики, работа которых реализуется медицинскими статистиками, врачами-статистиками и врачами-методистами. Достоверность, расчетные показатели и выводы о функционировании учреждений здравоохранения зависят от грамотности данных сотрудников [18-19]. Организация статистического учета подразумевает специфику региона и внедрение информационных технологий [20]. Однако исследования оценки эффективности организационной системы здравоохранения в большинстве случаев основаны на сведениях федерального статистического наблюдения без учета региональных особенностей учреждения здравоохранения. Данная проблема характеризуется отсутствием единых подходов при анализе и оценке эффективности деятельности управления ресурсным обеспечением компонентов региональных отраслевых организационных систем с учетом региональных особенностей [21].

Оптимальное распределение ресурсного обеспечения при использовании многолетней статистической информации может рассматриваться как одно из направлений повышения эффективности управления в региональных организационных системах, направленного на достижение заданных характеристик жизнедеятельности.

В работе [22] в целях моделирования взаимодействия объектов организационной системы с управляющим центром предлагается использовать рейтинговый подход, включающий в себя распределение рейтингового управления на три класса моделей: упорядочение, ресурсное обеспечение и рейтинговое состояние. При использовании полученных результатов, направленных на внутреннее и внешнее рейтинговое оценивание, происходит реализация многовариантности, когда принимаются управленческие решения.

В работах [23-25] в целях построения и развития цифровой среды управления в организационных системах рассматриваются и описываются формирование моделей и алгоритмов оптимизации распределения ресурсного обеспечения путем решения задачи моделирования процесса управления и алгоритмизации принятия управленческих решений при распределении ресурсного обеспечения.

В работе [26] рассматривается обоснование оптимизационного подхода к управлению ресурсным обеспечением в региональной отраслевой организационной системе, отличающейся процедурами интеграции результатов предиктивного анализа многолетней статистической информации в процесс принятия решений.

Принятие управленческих решений при распределении ресурсного обеспечения в региональных отраслевых организационных системах связаны с рядом особенностей.

1. Принятие решения по трем составляющим распределения ресурса:

– группам населения региона, которые могут выражаться в виде качественных (или атрибутивных: социальные группы населения, национальность, пол и др.) или количественных (которые можно подсчитать или измерить: возраст, уровень дохода и др.) группировочных признаков [27]. Метод группировки

является одним из важнейших методов, применяемых в статистическом учете при изучении массовых явлений для обобщения их данных [28];

- территориальным образованиям, в которых обеспечиваются ключевые условия жизнедеятельности населения [29], представляемые в виде районов региона, городских округов, субъектов РФ;
- временным периодам в рамках горизонта планирования – прогнозные значения будущих периодов.

Для первых двух составляющих управленческие решения определяются путем постановки и решения задач многоальтернативной оптимизации. Они позволяют определить перспективные подмножества групп населения и территориальных образований, для которых определенная экспертами потребность в дополнительном ресурсе даст в будущие периоды наибольший эффект.

Поскольку управленческие решения содержат наряду с формализованным выбором экспертную составляющую, они носят предварительный характер. Окончательное формализованное решение достигается при распределении предварительных оценок по временным интервалам с использованием оптимизационной модели динамического программирования.

2. Требования сбалансированности решений при распределении ресурсного обеспечения по трем составляющим с плановым объемом, заданным управляющим центром. Сбалансированные управленческие решения формируют необходимую основу стратегической устойчивости функционирования региональной отраслевой организационной системы и отражают достигнутое равновесие (компромисс) между средой и состоянием функционирующих в ней компонентов [30].

3. Учет множества показателей, фиксируемых региональной статистикой и характеризующих, с одной стороны, объемы использования ресурсного обеспечения (материально-технических, кадровых, финансовых), а с другой – показатели эффективности деятельности управляющего центра (качества и результативности труда, социальной и экономической эффективности) [31]. Статистический учет играет важную роль в управлении отраслевыми

региональными системами, поскольку позволяет осуществлять анализ многолетней статистической информации, необходимой для экспертного прогнозирования эффективности управленческих действий [32]. При формализованном описании процесса управления в региональных отраслевых организационных системах многолетняя статистическая информация представляется временными рядами значений показателей эффективности и объемов ресурсного обеспечения, зафиксированных в ежегодных отчетных документах [33]. Исходная информация, представленная временными рядами, дает возможность перейти к следующим этапам формализованного описания: моделированию [34] и оптимизации [26]. Эффективное объединение перечисленных этапов при принятии управленческих решений достигается применением средств интеллектуализации [35].

4. Преимущественная ориентация на экспертные оценки прошлого опыта управления ресурсным обеспечением при предсказании результатов функционирования системы в будущие временные периоды. При этом экспертные прогнозы обычно имеют нечисловой характер и нуждаются в дополнительной процедуре получения однозначно определенного числового значения прогнозируемого показателя [36], что характеризует ограниченность данного подхода. В работе [26] предлагается возможность преодоления этой ограниченности за счет формализованного представления зависимости функции интегрального эффекта от дополнительного ресурсного обеспечения, позволяющего перейти к поиску управленческих решений путем оптимизации.

5. Достаточность данных статистического учета для перехода от оценок по множеству показателей эффективности к оценке интегрального эффекта от использования ресурсного обеспечения. Требование достаточности реализуется путем выбора необходимого уровня детализации временных рядов. Степень детализации обуславливается целями и задачами использования данных статистики. Избыточная детализация затрудняет сбор, анализ и обработку статистической информации, увеличивая при этом расходы. Недостаточность статистических данных, в свою очередь, препятствует выявлению и анализу

существенных сторон и взаимосвязей процессов и явлений, происходящих в региональной отраслевой организационной системе. Следовательно, при составлении исходной статистической информации следует стремиться к оптимальному соотношению уровня детализации, точности и полноты исходных данных с учетом задач, для решения которых они используются [37].

1.2. Повышение эффективности управления за счет интеграции результатов предиктивного анализа и оптимизации

Целый ряд региональных отраслевых организационных систем связан с возможностью управляющего центра оценивать различные характеристики состояния и деятельности населения и влиять на их изменение в соответствии с установленными требованиями. При этом для принятия управленческих решений используются данные многолетней статистики, в рамках которой осуществляется периодическое оценивание характеристик системы.

К таким отраслевых региональным системам относится региональная система здравоохранения. Ее структура и структура многолетней медико-статистической информации имеют особенности, которые аналогичны и для других классов региональных отраслевых организационных систем [26]:

– структурная модель включает в себя субрегиональные подсистемы оказания медицинской помощи, что приводит к территориальной распределенности эпидемиологических и экономических показателей и определяет принятие решений управляющим центром на основе пространственно-временной информации;

– в структуре данных медицинской статистики [38] важную роль играют показатели, определяющие динамику заболеваемости как одного из направлений повышения эффективности оказания медицинской помощи, требующей ресурсного обеспечения.

Традиционно управляющий центр учитывает эти особенности отраслевой организационной системы с использованием эпидемиологического анализа [39]. Однако оценки уровня, частоты и динамики заболеваемости населения не интегрированы с экспертными оценками выбора наиболее эффективного варианта развития регионального здравоохранения [32, 40] по нескольким составляющим: нозологической, пространственной, временной. В большинстве случаев указанные экспертные решения не позволяют обеспечить рациональный баланс при распределении ресурсного обеспечения по перечисленным составляющим.

В [33] показана возможность перехода от экспертных методов принятия решения к формализованным на основе оптимизационных моделей. При этом подходе возникает необходимость построения математических моделей, учитывающих все составляющие с использованием многолетней медико-статистической информации. Такие возможности предоставляет одно из направлений применения искусственного интеллекта при переходе к формализованным методам принятия решений – предиктивная аналитика [41-42], которая характеризуется совокупностью методов автоматизированного анализа и интерпретации данных для прогнозирования результатов событий, происходящих в организационных системах здравоохранения, например, определение характера и частоты заболеваний, прогнозирование вероятности сопутствующих заболеваний у пациентов, определение интенсивности и распространения инфекционных заболеваний, классификация по уровню заболеваемости и многих других вариантов использования [43]. Сочетание предиктивной аналитики и экспертных управленческих решений в здравоохранении рассмотрено в [44].

В настоящее время для повышения эффективности управления в организационных системах здравоохранения активно используются различные методы моделирования [45-46], в том числе основанные на обработке многолетней

медицинско-статистической информации [39, 47]. В качестве исходной информации для построения моделей, ориентированных на оптимизацию ресурсного обеспечения медицинского обслуживания в регионе, используются временные ряды показателей заболеваемости и диспансеризации населения. При принятии управленческих решений ряд тенденций в динамике показателей временных рядов достаточно анализировать с использованием визуального моделирования [48-49]. В случае, когда определяется горизонт планирования ресурсного обеспечения, осуществляется прогностическое моделирование временных рядов с применением статистических методов [50] и методов машинного обучения [51].

Указанные формы моделирования позволяют при управлении ресурсным обеспечением в организационных системах здравоохранения более эффективно решать ряд задач:

- сравнение динамики показателей общей и первичной заболеваемости [52];
- установление адекватности процесса диспансеризации тенденциям в динамике заболеваемости [53];
- распределение ресурсного обеспечения с учетом оценки затрат по диспансеризации, проводимых региональными органами управления здравоохранением с целью снижения заболеваемости [54].

Статистические методы прогнозирования временных рядов в качестве зависимости будущего значения от прошлого используют некоторое уравнение. Наиболее распространенными и часто применяемыми статистическими методами прогнозирования временных рядов, которые можно отнести к классу методов функционального моделирования, являются следующие [55]:

- авторегрессионное интегрированное скользящее среднее (ARIMA). Модель применяется для анализа стационарных временных рядов на основе оценки линейной зависимости прогнозируемых значений от исторических;
- метод линейного Хольта (Holt linear). Представляет собой модель, в которой разделяется набор данных временных рядов на компоненты: тренд, сезонность и остаток;

– тройное экспоненциальное сглаживание (или метод Хольта-Винтерса, Holt Winter). Особенностью модели является применение экспоненциального сглаживания к компонентам в дополнение к уровню и тренду.

Далее рассмотрены роль и методы машинного обучения в предиктивном анализе. В последние годы применение методов машинного обучения в повседневной жизни стало распространенным явлением, чему способствовали активное развитие информационных технологий, стремительный рост накопленных объемов информации, появление новых математических методов и рост конкуренции в различных отраслях человеческой деятельности. Данные аспекты делают крайне востребованными подходы к использованию накопленных данных для повышения обоснованности и оперативности принятия управленческих решений [56].

Машинное обучение представляет собой совокупность методов искусственного интеллекта, направленных на разработку алгоритмов и статистических моделей, которые позволяют компьютерам обучаться и принимать решения на основе данных [57]. Алгоритмы машинного обучения позволяют автоматизировать процессы принятия решений путем обобщения известных примеров. Среди них наиболее востребовано контролируемое обучение (Supervised Learning), где пользователь предоставляет алгоритму пары объект-ответ, а алгоритм находит способ получения ответа по объекту. В частности, алгоритм способен выдать ответ для объекта, которого он никогда не видел раньше, без какой-либо помощи человека [58].

Особенностью прогнозирования временных рядов методами машинного обучения является формирование зависимости прогнозного значения от статистических данных, используемых при обучении, в виде некоторой структуры и правил перехода по ней [59]. Высокая степень точности прогноза, позволяющая машинам выполнять сложные задачи, повышается при обучении [60].

В качестве наиболее часто используемых методов машинного обучения при прогнозировании временных рядов выделяют моделирование с использованием деревьев решений (в т.ч. случайного леса – ансамбля, состоящего из нескольких

деревьев решений) [61-63] и нейросетевое моделирование [64-66]. В языке программирования Python прогностическое моделирование на основе методов машинного обучения осуществляется путем решения задачи регрессии, поэтому для обучения и прогнозов создаются, обучаются и используются регрессионные модели.

Цель задачи регрессии состоит в том, чтобы предсказать непрерывное числовое значение на основе одного или нескольких независимых признаков. Т.к. при работе с временными рядами важным аспектом является учет последовательности значений в ряде, возникает необходимость в дополнительных инструментах, при использовании которых регрессионные модели могли бы учитывать эту последовательность в качестве независимых переменных (признаков) для получения прогнозных значений. Одним из таких инструментов, реализованных в библиотеке Skforecast, является рекурсивное многошаговое прогнозирование [67], которое подразумевает использование прогнозируемых значений из предыдущих временных шагов в качестве входных данных для прогнозирования значений в последующих временных шагах. Принцип работы рекурсивного многошагового прогнозирования представлен на рисунке 1.1.

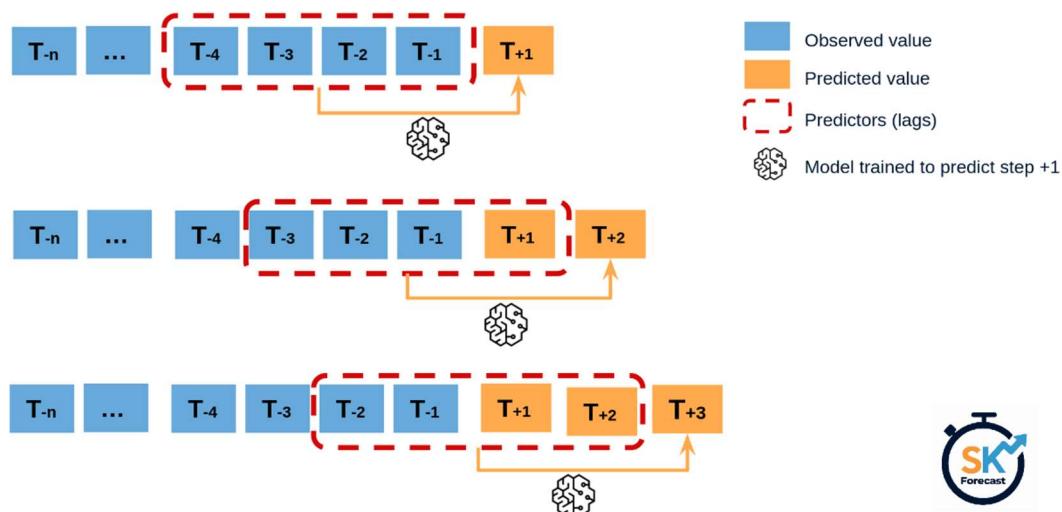


Рисунок 1.1 – Принцип работы рекурсивного многошагового прогнозирования

При оценке прогнозов временных рядов необходимо оценить качество ответов, выдаваемых статистическими методами или алгоритмами, основанными на машинном обучении. Наиболее распространенными показателями качества прогностических моделей являются: среднеквадратичная ошибка RMSE (Root Mean Squared Error), средняя абсолютная ошибка MAE (Mean Absolute Error) и коэффициент детерминации R². Совокупность нескольких показателей позволяет более комплексно оценить качество прогнозирования [68]. На языке программирования Python данные критерии качества разработаны в библиотеке Statmodels.

В системах машинного обучения также особое место занимает задача классификации, под которой подразумевают процесс упорядочения или распределения объектов по классам с целью отражения отношений между ними. Классификация – это логическая закономерность, позволяющая делать вывод относительно определения характеристик конкретной группы. Таким образом, для проведения классификации должны присутствовать признаки, характеризующие группу, к которой принадлежит то или иное событие или объект. Классификация относится к стратегии обучения с учителем, которое также называют контролируемым или управляемым обучением. Например, в региональных организационных системах здравоохранения задача классификации используется при прогнозировании появления побочных эффектов при лечении заболеваний [69], построении прогностических моделей для стратификации риска прогрессирования или возникновения рецидива болезней у пациентов [70] и др.

При распределении ресурсного обеспечения управляющим центром важно учитывать уровни заболеваемости в субрегиональной подсистеме оказания медицинской помощи, поскольку это позволяет помочь определиться с приоритетными компонентами региональной организационной системы, на которые необходимо обратить внимание в первую очередь в целях повышения эффективности оказания медицинской помощи. В целях определения степени заболеваемости, зафиксированной в субрегиональной организационной системе, можно выделить три уровня (класса) [71]:

- низкий – позволяет управляющему центру принять управленческое решение об уменьшении ресурсного обеспечения;
- средний – информирует управляющий центр о необходимости сохранения прежнего уровня ресурсного обеспечения или незначительного повышения на величину инфляции;
- высокий – отражает приоритетность субрегиональной организационной системы при распределении ресурсного обеспечения управляющим центром.

Таким образом, возникает задача многоклассовой классификации, в которой каждое наблюдение исходного набора данных необходимо отнести более чем к двум классам.

Выделим следующие наиболее распространенные базовые алгоритмы и методы классификации, способные работать с многоклассовой классификацией:

- k -ближайших соседей (k -nearest neighbors). Логика, лежащая в основе алгоритма, является одной из самых простых из всех алгоритмов машинного обучения. Суть метода заключается в вычислении расстояния от новой точки данных до всех других точек учебного набора данных;
- метод опорных векторов (SVM). Логика данного метода заключается в том, что при построении модели строится гиперплоскость или набор гиперплоскостей в пространстве высокой или бесконечной размерности, которые могут быть использованы для классификации;
- наивный Байес (NaiveBayes) – это вероятностная модель машинного обучения, которая используется для задачи классификации. Суть классификатора основана на теореме Байеса [72];
- дерево решений (DecisionTree) – представляет собой алгоритм, который моделирует решения в виде древовидной структуры;
- логистическая регрессия;
- нейронная сеть многослойный перцепtron (MLPClassifier). В основе данного алгоритма лежит обучение по методу обратного распространения ошибки (Backpropagation).

Помимо базовых алгоритмов и методов, многоклассовая классификация осуществляется также с использованием ансамблевых методов. Ансамблевое обучение – это тип обучения, в котором объединяются различные типы алгоритмов или один и тот же алгоритм несколько раз, чтобы сформировать модель прогнозирования, способную обеспечить более высокое качество классификации. Рассмотрим следующие ансамблевые методы при решении многоклассовой классификации:

- случайный лес (RandomForest) – метод объединяет несколько алгоритмов одного и того же типа, т.е. несколько деревьев решений, в результате чего образуется «лес деревьев»;
- бэггинг (bagging, bootstrap aggregating – бутстрэп-агрегирование) – представляет собой ансамблевый метод, который характеризуется тем, что подразумевает использование для каждого подклассификатора (слабого ученика) одного и того же алгоритма обучения, но обучение подклассификаторов происходит на разных случайных поднаборах обучающего набора [58];
- бустинг (Boosting, усиление) является ансамблевым методом, способным комбинировать несколько слабых учеников в одного сильного ученика [72]. Большинство методов бустинга предусматривает последовательное обучение подклассификаторов, причем каждый из них старается исправить ошибки своего предшественника. Наиболее распространенными видами данного ансамблевого метода являются Адаптивный бустинг (AdaBoost), Градиентный бустинг (GradientBoosting), CatBoost;
- стекинг (stacking, stacked generalization, стековое обобщение) основан на идеи использования нескольких, как правило, разнородных базовых классификаторов (т.е. комбинируются разные алгоритмы обучения), выдающих прогноз, который затем применяется в качестве признаков для обучения мета-классификатора. Основные отличия данного метода от бэггинга и бустинга заключаются в возможности использования разнородных слабых учеников, а также в применении мета-классификатора, способного объединить базовые

классификаторы, поэтому стекинг относят к «продвинутым» ансамблевым методам.

Оценка качества построенных тем или иным методом классификаторов является важнейшей компонентой жизненного цикла моделей, где эффект от ошибок 1-го и 2-го рода может генерировать существенные потери при принятии решений управляющим центром.

При этом часто возникает необходимость в изучении различных аспектов качества уже обученного классификатора. Будем считать, что классификатор имеет вид:

$$a(x) = [b(x) > t],$$

где $b(x)$ – прогноз классификатора, t – порог классификатора.

Для возможности оценки качества прогностических и классификационных моделей осуществляется разбиение исходных наборов данных на обучающие и тестовые выборки. В прогностическом моделировании в тестовую выборку обязательно включаются последние значения временного ряда. В классификационном моделировании тестовая выборка может формироваться случайным образом либо по какому-либо алгоритму. Таким образом, осуществляется предобучение в целях формирования группы моделей, оценка качества которых показала наилучший результат. После того как группа моделей была сформирована, для них осуществляется дообучение данными, которые составляли тестовую выборку. По итогу на основе обученных моделей формируются прогнозные значения, необходимые при управлении в региональной отраслевой организационной системе.

На практике чаще всего используют численную оценку качества алгоритмов [73]. Рассмотрим наиболее распространённые критерии к измерению качества классификационных моделей.

1. Доля правильных ответов.

Одним из критериев качества в задаче классификации является доля правильных ответов (accuracy):

$$accuracy(a, x) = \frac{1}{l} \sum_{i=1}^l [a(x_i) = y_i].$$

Тем не менее данный критерий имеет существенный недостаток. Если взять порог t меньше минимального значения прогноза $b(x)$ на выборке или больше максимального значения, то доля правильных ответов будет равна доле положительных и отрицательных ответов соответственно.

Таким образом, если в выборке 950 отрицательных и 50 положительных объектов, то при простом пороге $t = \max_i b(x_i)$ будет получена доля правильных ответов 0.95. Это означает, что доля положительных ответов сама по себе не несет никакой информации о качестве работы алгоритма $a(x)$ и вместе с ней следует анализировать соотношение классов в выборке. Также полезно вместе с долей правильных ответов вычислять базовую долю – долю правильных ответов алгоритма, всегда выдающего преобладающий класс.

Следует отметить, что при сравнении различных методов машинного обучения принято сообщать относительное уменьшение ошибки. Рассмотрим два алгоритма a_1 и a_2 с долями правильных ответов r_1 и r_2 соответственно, причем $r_1 > r_2$. Относительным уменьшением ошибки алгоритма a_2 называется величина

$$\frac{(1 - r_1) - (1 - r_2)}{1 - r_1}.$$

Если доля ошибок была уменьшена с 20% до 10%, то относительное улучшение составляет 50%. Если доля ошибок была уменьшена с 50% до 25%, то относительное улучшение также равно 50%, хотя данный прирост кажется более существенным. Если же доля ошибок была уменьшена с 0.1% до 0.01%, то относительное улучшение составляет 90%, что совершенно не соответствует действительности.

2. Матрица ошибок.

Другим распространенным критерием качества моделей классификации является матрица ошибок, которая позволяет оценить точность классификатора в

случае наличия несбалансированных классов, являясь, таким образом, более информативной.

Матрица ошибок представляет собой способ разбить объекты на четыре категории в зависимости от комбинации истинного ответа и ответа алгоритма (таблица 1.1).

Через элементы этой матрицы можно, например, выразить долю правильных ответов путем нахождения общей точности (OSR):

$$OSR = \frac{TP + TN}{TP + FP + FN + TN}.$$

Таблица 1.1 – Матрица ошибок

	$y = 1$	$y = -1$
$a(x) = 1$	True Positive (TP) количество истинно положительных наблюдений	False Positive (FP) количество ложноположительных наблюдений (ошибки II рода)
$a(x) = -1$	False Negative (FN) количество ложноотрицательных наблюдений (ошибки I рода)	True Negative (TN) количество истинно отрицательных наблюдений

Гораздо более информативными критериями являются показатели точности (Pr), полноты (чувствительности) (Se) и специфичности (Sp):

$$Pr = \frac{TP}{TP + FP},$$

$$Se = \frac{TP}{TP + FN},$$

$$Sp = \frac{TN}{TN + FN}.$$

Точность показывает, какая доля объектов, выделенных классификатором как положительные, действительно является положительными. Полнота

показывает, какая часть положительных объектов была выделена классификатором. Специфичность показывает долю верно классифицированных отрицательных объектов [74].

Важно отметить, что точность и полнота не зависят от соотношения размеров классов. Даже если объектов положительного класса на порядки меньше, чем объектов отрицательного класса, данные показатели будут корректно отражать качество работы алгоритма.

Существует несколько способов получить один критерий качества на основе точности и полноты. Один из них – показатель сбалансированной F-меры, гармоническое среднее точности и полноты:

$$F1 = \frac{2 \cdot Pr \cdot Se}{Pr + Se}.$$

Среднее гармоническое обладает важным свойством – оно близко к нулю, если хотя бы один из аргументов близок к нулю. Именно поэтому оно является более предпочтительным, чем среднее арифметическое (если алгоритм будет относить все объекты к положительному классу, то он будет иметь $Se = 1$ и $Pr \ll 1$, а их среднее арифметическое будет больше $1/2$, что недопустимо). Следует заметить, что геометрическое среднее также похоже на сглаженный вариант минимума, но при этом оно менее устойчиво к «выбросам» – например, для точности 0.9 и полноты 0.1 гармоническое среднее будет равно 0.18, а геометрическое 0.3.

3. Area Under Curve (AUC).

Широко используется такой интегральный критерий качества семейства, как площадь под ROC-кривой (Area Under ROC Curve, AUC-ROC) [75]. Рассмотрим двумерное пространство, одна из координат которого соответствует доле неверно принятых объектов (False Positive Rate, FPR), а другая – доле верно принятых объектов (True Positive Rate, TPR):

$$FPR = \frac{FP}{FP + TN},$$

$$Sp = \frac{TP}{TP + FN}.$$

Каждый возможный выбор порога t соответствует точке в этом пространстве. Всего различных порогов имеется $l + 1$. Максимальный порог $t_{max} = \max_i b(x_i)$ даст классификатор с $TPR = 0$, $FPR = 0$. Минимальный порог $t_{min} = \min_i b(x_i) - \varepsilon$ даст $TPR = 1$ и $FPR = 1$. ROC-кривая – это кривая с концами в точках $(0,0)$ и $(1,1)$, которая последовательно соединяет точки, соответствующие порогам $b(x_{(1)}) - \varepsilon, b(x_{(1)}), b(x_{(2)}), \dots, b(x_{(l)})$. Площадь под данной кривой называется AUC-ROC, и принимает значения от 0 до 1. Если порог t может быть подобран так, что алгоритм $a(x)$ не будет допускать ошибок, то AUC-ROC будет равен единице; если же $b(x)$ ранжирует объекты случайным образом, то AUC-ROC будет близок к 0.5.

Критерий AUC-ROC имеет большое число интерпретаций – например, он равен вероятности того, что случайно выбранный положительный объект окажется позже случайно выбранного отрицательного объекта в ранжированном списке, порожденном $b(x)$.

4. Валидационная точность.

Валидационная точность или скользящий контроль, или кросс-валидация (cross-validation, CV) – процедура эмпирического оценивания обобщающей способности алгоритмов, обучаемых по прецедентам. Суть данного критерия заключается в фиксировании некоторого множества разбиений исходной выборки на две выборки: обучающую и контрольную. Для каждого разбиения выполняется настройка алгоритма по обучающей выборке, затем оценивается его средняя ошибка на объектах контрольной выборки. Оценкой валидационной точности называется средняя по всем разбиениям величина ошибки на контрольных выборках. Если выборка независима, то средняя ошибка валидационной точности даёт несмешённую оценку вероятности ошибки. Это выгодно отличает её от средней ошибки на обучающей выборке, которая может оказаться смешённой (оптимистически заниженной) оценкой вероятности ошибки, что связано с явлением переобучения. Валидационная точность является стандартной методикой тестирования и сравнения алгоритмов классификации.

Базовые и ансамблевые методы классификации, а также критерии качества на языке программирования Python реализованы в библиотеке Scikit-Learn [76].

Управленческая деятельность неразрывно связана с принятием решений. Теорию принятия решений можно отнести к кибернетике и исследованию операций, а также к организационно-экономическому моделированию [77].

Принятие управленческих решений, основанных на многовариативном выборе, традиционно рассматривается как задача экспертного оценивания. В тех случаях, когда процесс выбора удается формализовать в виде зависимостей экстремальных и граничных требований от переменных, влияющих на принятие решений, используются методы оптимизации [34].

Принципы многоальтернативной агрегации определяют ряд особенностей методов моделирования задач оптимального синтеза сложных систем с варьируемой структурой. Они ориентируют на отражение содержательной постановки задач синтеза в виде оптимизационных моделей агрегации. В случае многоальтернативной агрегации сохраняется только структура этих моделей. Способ задания оптимизируемых переменных, на множестве которых строятся критерии и ограничения, подлежит разработке. Формализация многоальтернативного задания исходных элементов осуществляется путем введения вектора булевых переменных $x_m = (x_{m1}, \dots, x_{mn})$, $x_{mn} = \begin{cases} 1, & (m = \overline{1, M}, \\ 0, & n = \overline{1, N}), \end{cases}$ генерирование комбинаций которых индуцирует процесс дихотомической редукции разнообразия множества.

При машинном моделировании обращаются к реализациям случайных булевых переменных $\tilde{x}_{mn} (m = \overline{1, M}, n = \overline{1, N})$. Последнее равносильно проведению сложного конъюнктивного опыта A:

$$A = A_1 * A_2 \dots A_m \dots A_M,$$

где $A_m (m = \overline{1, M})$ – дизъюнктивные опыты. Когда редукция выполняется по нескольким ветвям, опыт A_m имеет $n = \overline{1, N_m}$ исходов. Формализация осуществляется путем введения булевых переменных с двумя индексами $x_{mn} =$

$\begin{cases} 1, & (m = \overline{1, M}, n = \overline{1, N_m}, \sum_{n=1}^{N_m} x_{mn} = 1), \\ 0, & \end{cases}$ которые при $N_m = 2 \forall m = \overline{1, M}$ вырождаются в переменные $x_m \begin{cases} 1 \\ 0 \end{cases} (m = \overline{1, M})$.

Переход к оптимизационным моделям, учитывающим многоальтернативность, состоит не только во введении булевых переменных. Для придания целенаправленного характера процессу ограничения разнообразия как оптимальному управлению структурой сложных систем, следует установить конструкцию аналитических моделей, отражающих экстремальные связи между показателями и переменными x_{mn} .

Дихотомической оптимизационной моделью называется математическая конструкция вида:

$$\begin{aligned} \Psi_{i1}(x_{mn}) &\rightarrow \text{extr}, i_1 \in I_1, \\ \Psi_{i2}(x_{mn}) &\leq b_{i2}, i_2 \in I_2, I_1 \cup I_2 = I, \\ \sum_{n=1}^{N_m} x_{mn} &= 1, m = \overline{1, M}, \\ x_{mn} &= \begin{cases} 1, & (m = \overline{1, M}, n = \overline{1, N}), \\ 0, & \end{cases} \end{aligned}$$

где I_1 – множество индексов показателей системы, требования к которым формализуется в виде критериев оптимизации; I_2 – множество индексов показателей системы, требования к которым формализуются в виде ограничений.

Представление обобщенной постановки задачи оптимального синтеза сложных систем с варьируемой структурой приводит к дихотомическим моделям большой размерности с разными по характеру булевыми переменными x_m и x_{mn} , и с большим числом ограничений разнотипной структуры.

Декомпозиция обобщенных постановок на локальные задачи в целях получения оптимизационных задач в совокупности дает решение проблемы агрегации в целом [78]. Поэтому первым этапом моделирования, основанного на принципах многоальтернативной агрегации, является этап декомпозиции обобщенной задачи синтеза. Другой особенностью формализации содержательной

постановки задачи синтеза, характеризующей варьируемую структуру сложной системы на уровне теоретико-множественных представлений, является сохранение количества выбора при переходе к дихотомическим оптимизационным моделям. Для этих целей необходимо выработать условия совместности, обеспечивающие адекватность исходным постановкам. Следовательно, очередной этап моделирования состоит в формировании условий адекватного перехода к аналитическим оптимизационным моделям.

Заключительным этапом моделирования является выбор структуры и размерности оптимизационных моделей, соответствующих локальным задачам синтеза, которые получены на этапе декомпозиции. При этом выбор следует проводить с учетом условий адекватности.

Особым типом неопределенностей является неопределенность в выборе цели при наличии множества критериев [79]. При решении оптимизационных задач многокритериальной оптимизации оптимизация векторного критерия связана с определением областей Парето, т.е. такого множества стратегий оптимизации, для которого невозможно одновременно улучшение всех скалярных критериев [80]. Выбор единственного решения из множества Парето осуществляется путем скаляризации вектора, заключающегося в выборе единственного критерия и преобразования остальных критериев в систему ограничений.

Другой метод скаляризации состоит в сворачивании отдельных критериев в единый глобальный критерий. При рассмотрении таких задач различают несколько подходов, например, априорное сворачивание, при котором исследователь на основе имеющейся информации может свернуть отдельные критерии в единый. В рамках такого подхода важным является выяснение возможных структур сворачивания критериев. Широкое распространение получили аддитивные свертки [81].

Апостериорное сворачивание осуществляется в ситуациях, когда имеющейся информации недостаточно для точной формулировки цели оптимизации. В данном случае возникает необходимость в активном накоплении информации путем проведения ограниченного числа экспериментов. Такие эксперименты состоят в

проверке справедливости некоторых гипотез о поведении исследуемой системы, выполнение которых приводит к частичному или полному устраниению неопределенности, связанной с выбором цели оптимизации. Данный подход можно назвать аксиоматическим.

Аксиоматический подход является с точки зрения эффективности наиболее приемлемым среди других подходов. При удачном использовании аксиом удается уменьшить, а часто и полностью устраниить неопределенность в выборе цели. Однако применение аксиоматических методов связано со значительными трудностями. Во-первых, имеющиеся системы аксиом не всегда являются легко проверяемыми в конкретных задачах. Во-вторых, построение систем аксиом затрудняется тем, что они должны быть, с одной стороны, несложными, а с другой стороны, достаточно информативными для уменьшения имеющейся неопределенности.

Возможен также адаптивный (стохастический) подход, при котором каждый из осуществляемых экспериментов не является столь информативным, как при аксиоматическом, и для устраниния начальной неопределенности в выборе цели требуется проведение значительного числа экспериментов. При таком подходе каждый последующий эксперимент осуществляется с учетом результата предыдущего и сворачивание критериев представляет лишь асимптотическую цель. Стохастический подход требует применения методов стохастической оптимизации, построенных с использованием правил локального улучшения. С этой целью осуществляется рандомизация множества критериев путем введения дискретной случайной величины, реализациями которой являются номера критериев. Задача состоит в настройке вероятностей привлечения критериев по информации от лица, принимающего решение, путем построения диалога.

В общем виде математическая модель принятия оптимального решения может быть сформулирована как задача нелинейного программирования. Эффективность численного решения задачи оптимизации во многом определяется ее размерностью. При одношаговом процессе принятия оптимального решения выбор численных значений для всех компонент вектора оптимизируемых

переменных осуществляется одновременно, что при большой размерности задачи может существенно затруднить поиск оптимального решения. Одним из подходов к преодолению отмеченной трудности является организация процесса принятия решения в виде N -шагового процесса принятия оптимального решения, в котором вектор оптимизируемых переменных разбивается на N не связанных между собой частей, называемых вектором управляемых параметров. Таким образом, данный подход позволяет вместо одного выбора вектора осуществить последовательность выборов векторов значительно меньше размерности, чем исходный вектор переменных.

Наряду с традиционными методами принятия решений на основе экспертной информации все большую значимость приобретают технологии интеллектуальной поддержки экспертных мнений с применением современных методов моделирования и оптимизации. Однако в большинстве случаев экспертная и оптимизационная составляющие функционируют как независимые подсистемы, при этом работы по интеграции предиктивной аналитики и методов оптимизации практически отсутствуют, поэтому возникает необходимость в разработке комплекса моделей и алгоритмов, позволяющих осуществить интеграцию результатов предиктивной аналитики ретроспективной информации, оптимизационного и экспертного моделирования в единый цикл принятия управлений решений в региональной отраслевой организационной системе.

1.3. Структурное моделирование региональной отраслевой организационной системы и интеллектуальной поддержки принятия решений при управлении ею

Показатели, фиксируемые в рамках ежегодного статистического учета, представляют собой две группы: показатели, характеризующие эффективность деятельности органов управления, и показатели, косвенно определяющие объемы

использованного ресурсного обеспечения. Данные группы характерны для любой региональной отраслевой организационной системы. Например, в региональной организационной системе здравоохранения к первой группе относятся показатели заболеваемости населения, а ко второй – в случае, если рассматривается деятельность, связанная с диспансеризацией населения, – количество лиц, находящихся на диспансерном учете. Затрачиваемое ресурсное обеспечение пропорционально объему диспансерного учета.

Одной из особенностей региональных отраслевых организационных систем является привязка перечисленных выше показателей, с одной стороны, к группам населения, объединенным по определенному признаку (в системе регионального здравоохранения к нозологическим группам заболеваемости), а с другой – к территориально распределенным субрегиональным образованиям (территориальным образованиям).

Введем следующие обозначения, необходимые для формирования структурной модели, характеризующей связи управляющего центра с распределением ресурсного обеспечения и показателями эффективности управления:

- $n = \overline{1, N}$ – нумерационное множество групп населения региона;
- $d = \overline{1, D}$ – нумерационное множество территориальных образований;
- $j = \overline{1, J}$ – нумерационное множество показателей эффективности деятельности;
- y_j – значения показателей эффективности;
- R – значения объемов определенного вида деятельности, характеризующие уровень ресурсного обеспечения;
- $y_j^o, j = \overline{1, J}$ – требования управляющего центра к показателям эффективности;
- R^o – плановый объем деятельности, характеризующий уровень ресурсного обеспечения и устанавливаемый управляющим центром;

– $t_1 = \overline{1, T_1}$ – временные (календарные) периоды фиксации данных статистического учета;

– $t = \overline{T_1 + 1, T_1 + T}$ – временные периоды горизонта планирования (прогнозирования) объемов деятельности, характеризующие уровень ресурсного обеспечения;

– $y_j(t_1), y_j(t)$ – временные ряды значений показателей соответственно для периодов $t_1 = \overline{1, T_1}, t = \overline{T_1 + 1, T_1 + T}$;

– $R(t_1), R(t)$ – временные ряды объемов деятельности соответственно для периодов $t_1 = \overline{1, T_1}, t = \overline{T_1 + 1, T_1 + T}$.

Структурная модель отражает действия управляющего центра:

- по определению требований $y_j^\circ, j = \overline{1, J}$ и планового объема R° ;
- по организации и использованию результатов статистического учета для групп населения и территориальных образований.

Схема структурной модели приведена на рисунке 1.2. В соответствии со структурной моделью действия управляющего центра в региональной отраслевой организационной системе, с одной стороны, направлены на установление планового ресурсного обеспечения для временного интервала $t - R^\circ(t)$ и требований к уровню показателей эффективности

$$y_j(t) \geq y_j^\circ, j = \overline{1, J}, \quad (1.1)$$

а с другой – на распределение $R^\circ(t)$ по следующим составляющим:

- группами населения, объединенными по некоторому признаку (в здравоохранении по определенным нозологиям), $n = \overline{1, N}$;
- субрегиональным образованиям $d = \overline{1, D}$, распределенными на территории региона;
- временными периодами, следующими за периодами многолетних статистических наблюдений $t_1 = \overline{1, T_1}, t = \overline{T_1 + 1, T_1 + T}$.

Будем считать результаты многолетней статистики

$$R^\circ(t_1), y_j(t_1), j = \overline{1, J}, t_1 = \overline{1, T_1}$$

и требования управляющего центра на будущие временные периоды

$$y_j^\circ(t), j = \overline{1, J}, t = \overline{T_1 + 1, T_1 + T}$$

$$R^\circ(t), t = \overline{T_1 + 1, T_1 + T}$$

исходными данными для формализации экстремальных и граничных требований при формировании оптимизационных моделей по перечисленным составляющим.

В качестве экстремального требования предлагается рассматривать требование максимизации эффекта от привлечения дополнительного ресурсного обеспечения $\Delta R(t)$, для каждого нового периода $t = \overline{T_1 + 1, T_1 + T}$ по сравнению с предыдущим периодом ($t - 1$):

$$\Delta R^\circ(t) = R^\circ(t) - R^\circ(t - 1)$$

при условии

$$R^\circ(t) > R^\circ(t - 1).$$

При этом достижение эффекта от влияния $R^\circ(t)$ определяется улучшением значений показателей эффективности

$$\Delta y_j(t) = y_j(t - 1) - y_j(t) > 0, j = \overline{1, J}, \quad (1.2)$$

а интегральный эффект является функцией локальных улучшений и дополнительного ресурсного обеспечения

$$Y(t) = F(\Delta R^\circ(t); \Delta y_j(t), j = \overline{1, J}; t). \quad (1.3)$$

Границное требование определяется необходимостью балансировать управленические решения на распределение ресурсного обеспечения с плановым уровнем $R^\circ(t)$.

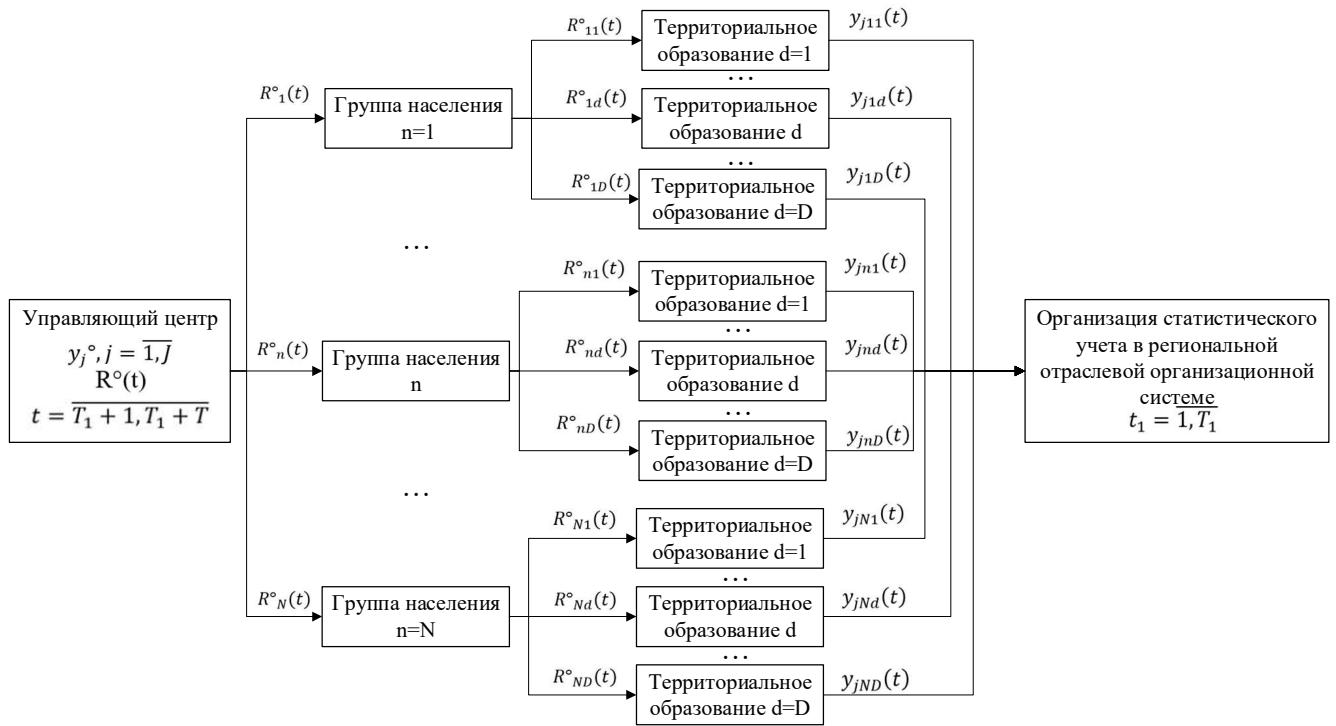


Рисунок 1.2 – Структурная модель взаимодействия управ器яющего центра с компонентами региональной отраслевой организационной системы при управлении распределением ресурсного обеспечения

Структурная модель управления распределением ресурсного обеспечения базируется на двухуровневой системе поддержки принятия решений:

- использование методов искусственного интеллекта для получения результатов предиктивной аналитики путем формирования прогностических моделей $\Delta R(t), \Delta y_j(t), j = \overline{1, J}, Y(t), t = \overline{T_1 + 1, T_1 + T}$;
- интеграция результатов предиктивной аналитики в процесс принятия управленческих решений с использованием интеллектуальной оптимизации.

Структурная модель интеллектуализации управления распределением ресурсного обеспечения в региональной отраслевой организационной системе на основе ретроспективной информации представлена на рисунке 1.3.



Рисунок 1.3 – Структурная модель интеллектуализации управления распределением ресурсного обеспечения в региональной отраслевой организационной системе на основе ретроспективной информации

Выводы первой главы

1. Для характеристики региональных отраслевых организационных систем как отдельного класса объектов управления, использующих в качестве информационного обеспечения данные ежегодного статистического учета, следует определить те особенности, которые влияют на разнообразие процедур принятия решений. К таким особенностям прежде всего относится наличие в ретроспективной информации данных о показателях эффективности и объема ресурсного обеспечения, детализированные по группам населения и территориальным образованиям региона.

2. Интеграция результатов предиктивного анализа и оптимизационного подхода способствует повышению эффективности управления в региональных отраслевых организационных системах с использованием ретроспективной информации за счет возможностей экспертной трансформации визуальных и функциональных представлений прогностических моделей при формировании формализованных описаний, экстремальных и граничных требований оптимизационных задач от оптимизируемых переменных.

3. Базовой основой перехода от экспертной формы принятия решения к оптимизации являются структурные модели как самой региональной отраслевой организационной системы, так и системы управления ею. Именно структурное моделирование позволяет организовать поддержку экспертных решений на основе двухуровневой системы и ее реализацию в алгоритмической схеме объединения всех этапов предиктивного анализа, ориентированного на оптимизационный подход.

ГЛАВА 2. ФОРМИРОВАНИЕ РЕЗУЛЬТАТОВ ПРЕДИКТИВНОГО АНАЛИЗА РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ НА ОСНОВЕ ОБЪЕДИНЕНИЯ ВИЗУАЛЬНОГО МОДЕЛИРОВАНИЯ И МОДЕЛИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

В п. 1.2 показана возможность повышения эффективности управления в региональной отраслевой организационной системе на основе использования результатов предиктивного анализа ретроспективной информации при принятии управленческих решений путем моделирования с использованием средств интеллектуализации принятия решений.

По данным статистического учета формируются временные ряды нормированных значений показателей эффективности и ресурсного обеспечения. Для последующей оптимизации важно, что эта информация детализована по группам населения и территориальным образованиям региональной отраслевой организационной системы.

Визуализация графических представлений временных рядов позволяет использовать механизмы наглядно-образной интуиции для экспертного анализа особенностей динамических процессов. В этом случае возможность визуального моделирования тенденций и темпов изменения показателей эффективности, степени адекватности изменения ресурсного обеспечения изменениям показателей эффективности для предварительной оценки того массива данных, который наиболее характерен в качестве тестовой выборки.

Результаты предиктивного анализа заключаются в построении и использовании прогностических моделей временных рядов. На основе тестовой выборки формируются разные классы моделей путем машинного обучения. Наилучшие по точности объединяются в ансамбль предобученных

прогностических моделей. Кроме того, необходимость оптимизации по величине интегрального эффекта от использования дополнительного ресурсного обеспечения обуславливает построение единой прогностической модели в рамках всего горизонта прогнозирования.

В процессе принятия управленческих решений используются результаты предиктивного анализа на выборках, отличающихся от тестовой. В этом случае осуществляется дообучение предобученных прогностических моделей, объединенных в ансамбль. Перечисленные действия создают предпосылки для эффективного моделирования с использованием средств интеллектуализации принятия решений при управлении в региональной отраслевой организационной системе.

2.1. Структура объединенной алгоритмической схемы визуального моделирования и моделирования с использованием машинного обучения

Рассмотрим процесс формирования результатов предиктивного анализа для их использования при принятии управленческих решений в региональной отраслевой организационной системе как многоэтапную алгоритмическую процедуру, объединяющую визуальное моделирование и моделирование на основе машинного обучения с целью последующей оптимизации.

Исходная информация, основанная на данных многолетнего статистического учета, структурируется в массивы значений показателей эффективности и объемов ресурсного обеспечения, детализованных по группам населения, территориальным образованиям и временным периодам. Как показано в п. 1.1, для управляющего центра важны преобразованные массивы данных:

– нормированные на определенное количество населения значения показателей эффективности

$$\begin{aligned}
 y_j(t), j = \overline{1, J}, t = \overline{1, T_1}; \\
 y_{jn}(t), j = \overline{1, J}, n = \overline{1, N}, t = \overline{1, T_1}; \\
 y_{jnd}(t), j = \overline{1, J}, n = \overline{1, N}, d = \overline{1, D}, t = \overline{1, T_1};
 \end{aligned} \tag{2.1}$$

– нормированные на определенное количество населения значения дополнительного ресурсного обеспечения

$$\begin{aligned}
 \Delta R(t), t = \overline{1, T_1}; \\
 \Delta R_n(t), n = \overline{1, N}, t = \overline{1, T_1}; \\
 \Delta R_{nd}(t), n = \overline{1, N}, d = \overline{1, D}, t = \overline{1, T_1};
 \end{aligned} \tag{2.2}$$

– значения интегрального эффекта от использования дополнительного ресурсного обеспечения

$$\begin{aligned}
 Y(t) = F \left(\Delta R(t), \Delta y_j(t) \right), t = \overline{1, T_1}; \\
 Y_n(t) = F_n \left(\Delta R_n(t), \Delta y_{jn}(t) \right), n = \overline{1, N}, t = \overline{1, T_1}; \\
 Y_{nd}(t) = F_{nd} \left(\Delta R_{nd}(t), \Delta y_{jnd}(t) \right), n = \overline{1, N}, d = \overline{1, D}, t = \overline{1, T_1};
 \end{aligned} \tag{2.3}$$

где одним из вариантов функции F может служить отношение величины улучшения показателей эффективности к величине дополнительного ресурсного обеспечения

$$\begin{aligned}
 F \left(\frac{\Delta y_j(t)}{\Delta R(t)} \right), j = \overline{1, J}; \\
 F_n \left(\frac{\Delta y_{jn}(t)}{\Delta R_n(t)} \right), n = \overline{1, N}, j = \overline{1, J}; \\
 F_{nd} \left(\frac{\Delta y_{jnd}(t)}{\Delta R_{nd}(t)} \right), n = \overline{1, N}, d = \overline{1, D}, j = \overline{1, J}.
 \end{aligned} \tag{2.4}$$

Все массивы данных (2.1)-(2.3) представляют собой временные ряды, графические представления которых позволяют сформировать экспертные мнения об особенностях динамики на основе наглядно-образных механизмов интуиции. Поэтому предлагается включить в алгоритмическую схему возможность

использования предварительного визуального моделирования для последующего предиктивного анализа, включающего прогностическое моделирование временных рядов (2.1)-(2.3) с учетом вычисления функции интегрального эффекта на основе преобразований (2.4).

Визуальное моделирование включает в себя ряд сравнительных оценок экспертов управляющего центра региональной отраслевой организационной системы:

- тенденций изменений $\Delta y_{jn}(t)$ для разных групп населения $n = \overline{1, N}$ и разных показателей $j = \overline{1, J}$ за временные периоды $t = \overline{1, T_1}$;
- темпов изменения (интенсивности) $\Delta y_j(t)$, $\Delta y_{jn}(t)$, $\Delta y_{jn}(t)$;
- адекватности изменения дополнительного ресурсного обеспечения $\Delta R(t)$, $\Delta R_n(t)$, $\Delta R_{nd}(t)$ для улучшения значения показателей $\Delta y_j(t)$, $\Delta y_{jn}(t)$, $\Delta y_{jnd}(t)$.

Эти результаты являются основой для экспертного обоснования одного из массивов (2.2), соответствующих $n = n_1$, $d = d_1$ в качестве тестовой выборки, в которой отражаются характерные тенденции изменений $\Delta y_{jn1}(t)$, темпы этих изменений $\Delta y_{jn1}(t)$, $\Delta y_{jn1d1}(t)$, приемлемая степень адекватности изменений $\Delta R_{n1}(t)$, $\Delta R_{n1d1}(t)$ и $\Delta y_{jn1}(t)$, $\Delta y_{jn1d1}(t)$.

Прогностическое моделирование на основе предиктивного анализа направлено на формирование ансамбля предобученных моделей с использованием тестовой выборки. Для этого осуществляется машинное обучение разных классов моделей для прогнозирования временных рядов (2.1), (2.2) на временные периоды $t = \overline{T_1 + 1, T_1 + T}$: функциональных, нейросетевого, на основе деревьев решений и случайного леса. Обученные модели сравниваются по точности прогнозирования. В результате получаем ансамбль моделей, наилучших по точности для каждого j -го показателя, n_1 -й группы населения, d_1 -го территориального образования региональной отраслевой организационной системы. С целью перехода к использованию результатов предиктивного анализа при принятии управленческих решений на основе моделирования с использованием средств интеллектуализации

осуществляется дообучение моделей, полученных с использованием тестовой выборки для всех групп населения и территориальных образований.

Важную роль для оптимизации распределения дополнительного ресурсного обеспечения между временными периодами $t = \overline{T_1 + 1, T_1 + T}$ играют модели интегрального эффекта, обученные на основе массивов (2.3) и преобразований (2.4). В этом случае предлагается проводить сравнительный анализ разных моделей, обученных на основе (2.3):

$$Y(t) = F(\Delta R(t), t), t = \overline{T_1 + 1, T_1 + T}$$

с моделями, в которых выборка для обучения формируется на основе вычисления (2.4) по прогностическим моделям на стадии предиктивного анализа.

Структура объединенной алгоритмической схемы визуального моделирования и моделирования с использованием машинного обучения приведена на рисунке 2.1.

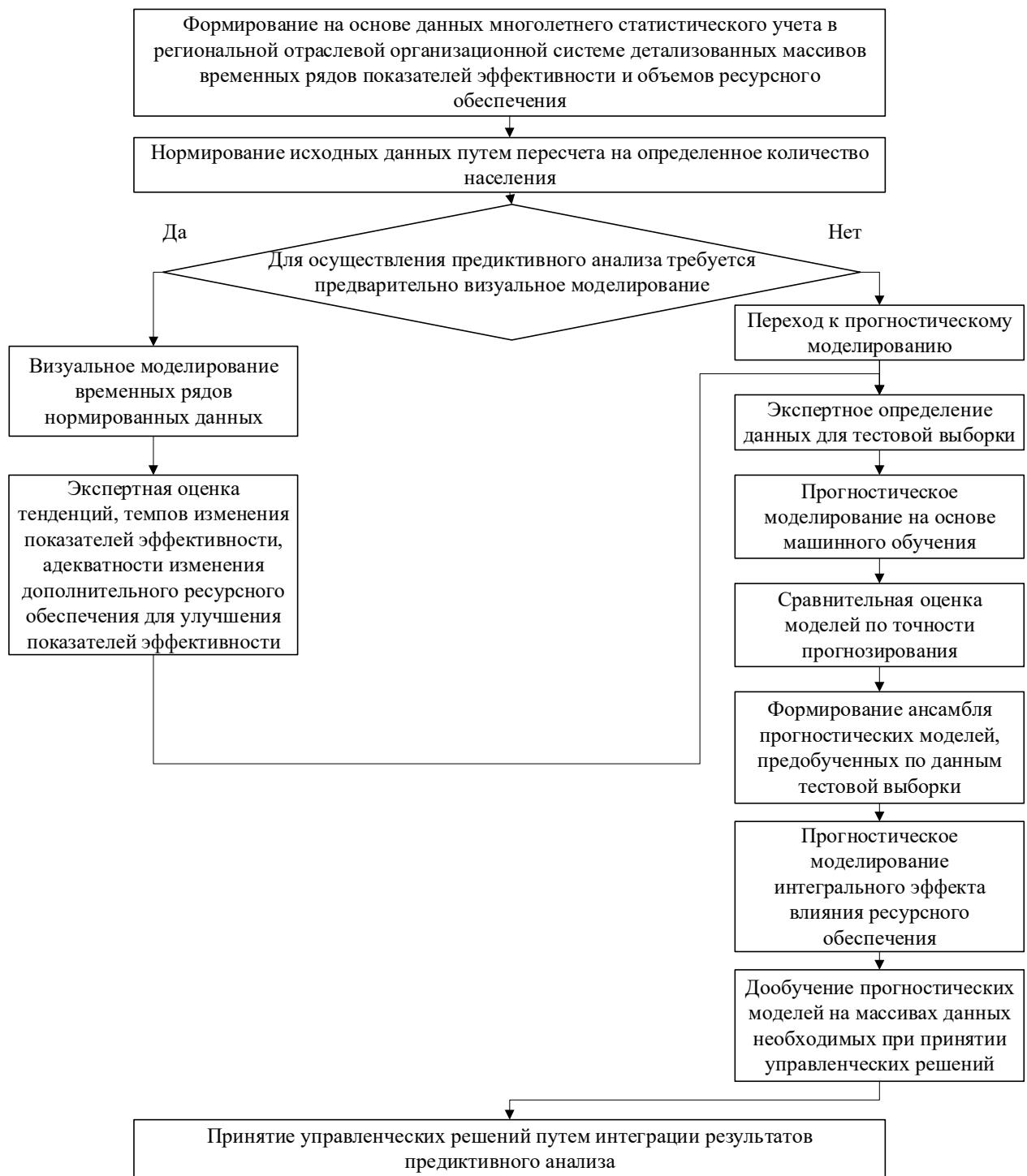


Рисунок 2.1 – Структура объединенной алгоритмической схемы визуального моделирования и моделирования с использованием машинного обучения

2.2. Формирование ансамбля прогностических моделей показателей эффективности и объемов дополнительного ресурсного обучения

Формирование ансамбля прогностических моделей характерно для разных отраслей региональных организационных систем (здравоохранение, образование, социальное обеспечение) и осуществляется идентично. В качестве примера была выбрана региональная организационная система здравоохранения, для которой был сформирован ансамбль моделей на основе прогностического моделирования по группе заболеваемости артериальная гипертензия, в частности, по следующим двум нозологиям: болезни, характеризующиеся повышенным кровяным давлением и эссенциальная гипертензия для лиц старших возрастных групп и их диспансеризации на основе многолетней медико-статистической информации.

Для анализа заболеваемости сердечно-сосудистой системы среди населения Воронежской области исследовались показатели на основе официальных данных статистики за десятилетний период с 2013 по 2022 гг.

Исследовались следующие показатели:

- общая заболеваемость;
- первичная заболеваемость;
- число больных с впервые в жизни установленным диагнозом, взятых под диспансерное наблюдение;
- число больных с впервые в жизни установленным диагнозом, выявленных при профосмотре;
- число больных с впервые в жизни установленным диагнозом, выявленных при диспансеризации;
- число лиц, состоящих под диспансерным наблюдением на конец года.

В таблице 2.1 представлен фрагмент многолетней статистики уровня общей заболеваемости болезнями, характеризующимися повышенным кровяным

давлением среди населения Воронежской области, нормированные на 100 000 населения старше трудоспособного возраста.

Таблица 2.1 – Фрагмент многолетней статистики уровня общей заболеваемости болезнями, характеризующиеся повышенным кровяным давлением среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

Район/Год	2013	2014	...	2020	2021	2022
Аннинский	11792,2	15480	...	32903,2	45673,4	45137,2
Бобровский	29318,4	23983	...	41032,2	39498,8	45141,5
Богучарский	23294,2	28160,5	...	42713,9	39798	48394,3
Борисоглебский	28871,8	27923,2	...	35587,4	44360,5	42290,5
Бутурлиновский	20068,7	21299,4	...	49524,2	45548	44348,8
Верхнемамонский	17334,3	6857,6	...	22480,1	22598,8	34272,1
...
Россошанский	20168	20061,5	...	32403,7	34949,6	40602,7
Семилукский	10380,9	12916,4	...	27226,2	30111,3	49777,1
Таловский	26327,5	25799,6	...	34910,3	29029	41447,7
Терновский	12874,1	12944,4	...	19519	20361,9	33987,2
Хохольский	28301,5	28257,7	...	40560,2	38703,9	40633,9
Эртильский	3972,3	8183,5	...	14066,1	13179,5	13509,2
ГО г. Воронеж	24429,7	27042,7	...	43906,8	42929,6	50204,5
Всего по области	20735,1	21856,4	...	38331,7	38136,1	43421,3

Субрегиональные организационные системы (районы Воронежской области) разделены на три уровня заболеваемости, описанные в п. 1.2, путем решения задачи многоклассовой классификации.

На основе имеющейся ретроспективной информации был подготовлен исходный обучающий набор данных, состоящий из следующих атрибутов:

- Район – характеризуется как субрегиональная организационная система, представленная в виде районов Воронежской области (тип данных *object*);
- Год – период временного ряда многолетней медико-статистической информации (тип данных *int64*);
- Показатель11 – данные общей заболеваемости, болезни, характеризующиеся повышенным кровяным давлением (тип данных *float64*);

- Показатель12 – данные первичной заболеваемости, болезни, характеризующиеся повышенным кровяным давлением (тип данных float64);
- Показатель13 – данные по числу больных с впервые в жизни установленным диагнозом, взятых под диспансерное наблюдение, болезни, характеризующиеся повышенным кровяным давлением (тип данных float64);
- Показатель14 – данные по числу больных с впервые в жизни установленным диагнозом, выявленных при профосмотре, болезни, характеризующиеся повышенным кровяным давлением (тип данных float64);
- Показатель15 – данные по числу больных с впервые в жизни установленным диагнозом, выявленных при диспансеризации, болезни, характеризующиеся повышенным кровяным давлением (тип данных float64);
- Показатель16 – данные по числу лиц, состоящих под диспансерным наблюдением на конец года, болезни, характеризующиеся повышенным кровяным давлением (тип данных float64);
- Показатель21 – данные общей заболеваемости, эссенциальная гипертензия (тип данных float64);
- Показатель22 – данные первичной заболеваемости, эссенциальная гипертензия (тип данных float64);
- Показатель23 – данные по числу больных с впервые в жизни установленным диагнозом, взятых под диспансерное наблюдение, эссенциальная гипертензия (тип данных float64);
- Показатель24 – данные по числу больных с впервые в жизни установленным диагнозом, выявленных при профосмотре, эссенциальная гипертензия (тип данных float64);
- Показатель25 – данные по числу больных с впервые в жизни установленным диагнозом, выявленных при диспансеризации, эссенциальная гипертензия (тип данных float64);
- Показатель26 – данные по числу лиц, состоящих под диспансерным наблюдением на конец года, эссенциальная гипертензия (тип данных float64);

– Класс – уровень заболеваемости (низкий – 1, средний – 2, высокий – 3) на основе данных 6 показателей (тип данных int64) – целевое значение, которое необходимо спрогнозировать.

Полученный исходный набор данных содержит 320 строк. При маркировке классов в исходном обучающем наборе данных была использована экспертная оценка уровней заболеваемости и диспансеризации по среднегодовому значению каждого из показателей и затем сформирована в виде итогового класса на основе подытоговых классов в разрезе нозологии каждого из шести показателей путем расчета их среднего арифметического и округлением до целого.

Перед началом проведения обучения классификационных моделей необходимо провести предварительную подготовку данных, которая требуется для разработки хорошей модели машинного обучения и нужна для совместимости с требованиями алгоритмов и методов классификации данных. При исследовании наборов данных активно применяются технологии визуализации данных, которые стали неотъемлемой частью современной науки, т.к. они позволяют обеспечивать представление данных в виде, наиболее обеспечивающем результативную работу пользователей [82]. Для визуализации данных в целях исследования состояния атрибутов набора данных использовался Python. Технологии и библиотеки визуализации данных на данном языке программирования рассмотрены в [83]. Важным этапом при подготовке данных к обучению является обнаружение и обработка отсутствующих данных в обучающем наборе данных.

Построена гистограмма пропущенных данных в целях определения количества строк с отсутствующими значениями (рисунок 2.2).

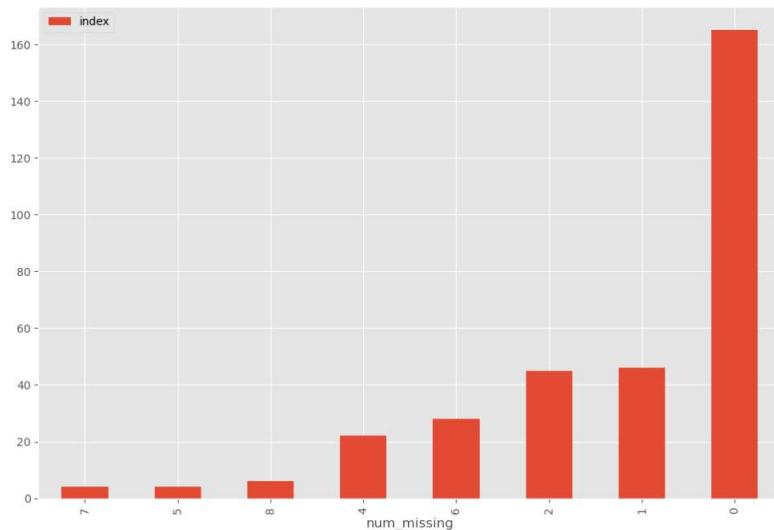


Рисунок 2.2 – Гистограмма отсутствующих значений в наборе данных

Гистограмма позволяет обнаружить, что в исходном наборе данных около 45 записей имеют пропуски по одному или двум атрибутам, около 30 записей имеет 6 пропусков, около 20 записей – 4 пропуска и около 5 строк с пропусками в 5, 7 и 8 атрибутах, при этом около 165 строк не имеют никаких пропусков. Процентное соотношение пропущенных значений представлено в таблице 2.2.

Далее сформируем тепловую карту пропущенных значений, представленную на рисунке 2.3. Недостаточность значений характеризуются отсутствием данных за 2013 год по районам в атрибутах «Показатель13», «Показатель23», а в атрибутах «Показатель14», «Показатель15», «Показатель24», «Показатель25» – за 2013 и 2014 гг. Их можно отнести к информативным, поэтому для заполнения пропусков их было принято решение заполнить медианными значениями в разрезе атрибута «Район». Пропуски в атрибутах «Показатель21», «Показатель22» были заполнены также медианными значениями в разрезе атрибута «Район», а для случаев отсутствия данных по какому-либо району – медианными значениями по соответствующему столбцу в разрезе атрибутов «Класс» и «Год», поскольку данные строки содержат информативные значения по большинству остальных атрибутов, а усреднение позволит сгладить влияние пропусков при обучении моделей и является наиболее быстрым в реализации.

Таблица 2.2. Соотношение пропущенных значений в наборе данных

Атрибут	Процент пропущенных значений, %
Район	0
Год	0
Показатель11	0
Показатель12	0
Показатель13	10
Показатель14	20
Показатель15	20
Показатель16	0
Показатель21	18
Показатель22	35
Показатель23	10
Показатель24	20
Показатель25	20
Показатель26	0
Класс	0

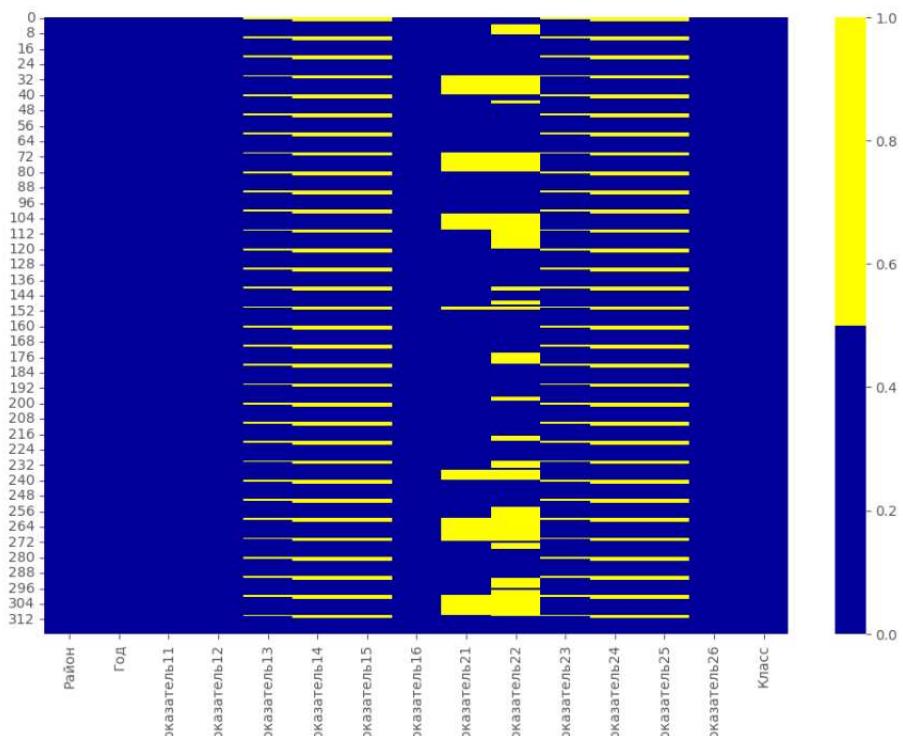


Рисунок 2.3 – Тепловая карта пропущенных значений в наборе данных

Следующим этапом подготовки данных является обработка категориальных признаков. Большинство методов классификации требуют для корректной работы численные признаки для проведения обучения модели, поэтому в исследуемом обучающем наборе следует преобразовать все категориальные признаки в числовые [84]. В наборе данных присутствует категориальный атрибут «Район». Он относится номинальному виду, что означает невозможность проранжирования его в целях упорядочивания, поэтому данный атрибут был преобразован в набор числовых признаков с использованием метода Binary Encoding («Двоичное кодирование»), который предполагает преобразование каждой категории в двоичный код и представление его в виде последовательности двоичных цифр (нулей и единиц) [85] и позволяет не увеличивать существенно объем набора данных, как происходит при использовании метода One-Hot Encoder за счет значительного числа исследуемых районов.

Следующим этапом предварительной подготовки данных является нормализация, включающая преобразование атрибутов в близкие масштабы для повышения производительности и стабильности алгоритмов и методов классификации машинного обучения [86]. Было принято решение осуществить нормализацию с использованием метода Robust Scaling («Надежное масштабирование»), который использует медиану и межквартильный размах для масштабирования признаков, что делает его надежным к выбросам [87].

Перед началом проведения обучения классификационных моделей необходимо проанализировать соотношение классов относительно друг друга. Несбалансированные обучающие наборы данных влияют на производительность моделей классификации, поэтому для балансировки классов используется метод синтетической выборки меньшинства (SMOTE), которые решают проблему дисбаланса классов, генерируя синтетические (подобные) данные для класса меньшинства [88]. На рисунке 2.4 представлена гистограмма соотношения классов.

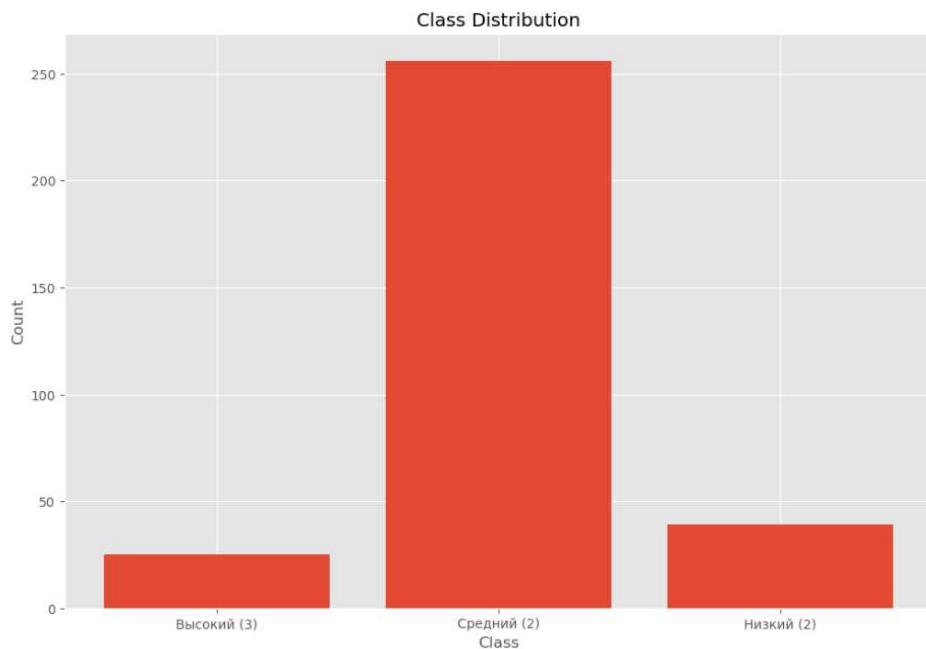


Рисунок 2.4 – Гистограмма соотношения классов до балансировки

Видно, что средний уровень (2) имеет значительно больше строк в обучающем наборе данных относительно остальных, поэтому необходимо произвести балансировку данных методом SMOTE. После проведения балансировки соотношение классов в наборе данных теперь выглядит следующим образом (рисунок 2.5). При этом количество записей в обучающем наборе данных увеличилось до 767 строк.



Рисунок 2.5 – Гистограмма соотношения классов после балансировки

Таким образом, было получено распределение классов, представленное на рисунке 2.6.



Рисунок 2.6 – Распределение классов после балансировки

Для проведения классификации были использованы следующие базовые и ансамблевые методы, поддерживающие многоклассовую классификацию, описанные в п.1.2:

- SVM – базовый метод опорных векторов;
- KNeighbors – базовый метод k -ближайших соседей;
- NaiveBayes – базовый метод наивный Байес;
- DecisionTree – базовый алгоритм дерево решений;
- MLPClassifier – базовый алгоритм нейронной сети многослойный перцептрон;
- AdaBoost – ансамблевый метод адаптивный бустинг;
- CatBoost – ансамблевый метод одной из схем градиентного бустинга;
- GradientBoosting – ансамблевый метод градиентный бустинг;

- RandomForest – ансамбль случайный лес;
- Bagging – ансамблевый метод бэггинг;
- Stacking – ансамблевый метод стекинг.

Для оценки качества обучения классификационных моделей использовались метрики, представленные в таблице 2.3. При обучении моделей классификации набор данных был разделен на обучающую и тестовую выборки в соотношении 80/20. Подбор параметров классификационных моделей осуществлялся с использованием метода поиска по сетке [89].

Таблица 2.3 – Критерии качества классификационных моделей

Название	Описание
accuracy	Общая точность
validation accuracy	Валидационная точность (кросс-валидация)
recall	Чувствительность
precision	Специфичность
F1	Показатель F1-меры
ROC AUC	Показатель ROC AUC
Confusion Matrix	Матрица ошибок (ошибки первого (Err 1) и второго (Err 2) рода)

В таблице 2.4 приведены результаты обучения классификационных моделей и оценка качества классификации.

Таблица 2.4 – Оценка качества классификаторов

	accuracy, %	validation accuracy, %	recall, %	precision, %	F1, %	ROC AUC, %	Err I	Err II	Всего ош-к
KNeighbors	87.01	84.69	87.01	88.16	86.44	87.01	11	9	20
SVM	90.26	88.92	90.26	90.25	90.15	90.26	8	7	15
NaiveBayes	81.82	80.94	81.82	83.08	81.27	81.82	26	2	28
DecisionTree	96.75	94.47	96.75	96.79	96.73	96.75	3	2	5
MLPClassifier	88.28	81.4	88.28	88.35	88.15	88.28	6	9	15
AdaBoost	92.21	86.32	92.21	92.24	92.22	92.21	5	7	12
CatBoost	98.05	95.93	98.05	98.11	98.04	98.05	2	1	3
GradientBoosting	97.4	94.3	97.4	97.52	97.39	97.4	3	1	4
RandomForest	98.05	97.07	98.05	98.09	98.05	98.05	3	0	3
Bagging	98.05	96.09	98.05	98.09	98.05	98.05	3	0	3
Stacking	98.7	96.79	98.7	98.71	98.7	98.7	2	0	2

Среди базовых методов классификации наилучший результат по качеству был получен у дерева решений. Наихудший результат – у метода Наивный байес. Среди ансамблевых методов наилучший результат оказался у Stacking, который был достигнут благодаря использованию в качестве базовых классификаторов базового метода дерева решений и ансамбля градиентный бустинг, показавших хороший результат при отдельном обучении (таблица 2.4), а в качестве метаклассификатора был использован ансамбль CatBoost, показавший один из наилучших результатов среди ансамблей. При обучении ансамблевого метода Bagging наилучший результат был получен при использовании в качестве подклассификатора ансамбля случайный лес, но оценка качества показала, что при использовании нескольких ансамблей случайного леса, итог оказался подобным тому, что был получен при обучении этой классификационной модели отдельно. Результаты качества обучения ансамбля CatBoost показали незначительное улучшение по сравнению с ансамблем GradientBoosting (на одну ошибку меньше), что в рамках данной задачи подтверждает улучшенный алгоритм построения классификационных моделей машинного обучения относительно основы, на которой он построен.

По итогам 2022 года районы Воронежской области распределены по следующим группам, соответствующих уровням по степени заболеваемости на основе показателей по двум нозологиям:

- группа 1 – Эртильский район;
- группа 2 – Аннинский, Борисоглебский, Верхнемамонский, Верхнекавский, Воробьевский, Калачеевский, Каменский, Кантемировский, Каширский, Нижнедевицкий, Новоусманский, Новохоперский, Павловский, Панинский, Поворинский, Подгоренский, Репьевский, Россонанский, Семилукский, Таловский, Терновский, Хохольский районы;
- группа 3 – Бобровский, Богучарский, Бутурлиновский, Грибановский, Лискинский, Ольховатский, Острогожский, Петропавловский, Рамонский районы.

При проведении прогностического моделирования наборы данных были разделены на обучающую и тестовую выборки. Обучающие данные содержат информацию до 2020 года наблюдений, а тестовая выборка включает в себя 2021-2022 годы наблюдений.

Сравнивались результаты прогнозирования с использованием следующих методов:

1. Функциональное моделирование:

– авторегрессионное интегрированное скользящее среднее (ARIMA);

– метод линейного Хольта (Holt linear, HL);

– тройное экспоненциальное сглаживание (или метод Хольта-Винтерса, Holt Winter, HW).

2. Моделирование с использованием деревьев решений (Decision tree, DT) и случайного леса (Random forest, RF).

3. Нейросетевое моделирование с использованием многослойного перцептрона (Multi-layer Perceptron, MLP).

Методы машинного обучения DT, RF и MLP применены рамках метода рекурсивного многошагового прогнозирования для получения прогнозных оценок в используемых регрессионных моделях на будущие периоды. Ошибка прогнозирования каждого из используемых методов оценивалась с использованием показателей среднеквадратичной ошибки RMSE, средней абсолютной ошибки MAE и коэффициента детерминации R2. В таблицах 2.5-2.6 приведены сравнительные результаты прогностического моделирования временных рядов по данным общей заболеваемости и диспансеризации по нозологии болезней, характеризующихся повышенным кровяным давлением. Результаты по остальным временным рядам приведены в приложении В. Серым цветом отмечены методы, в которых зафиксировано наилучшее качество прогнозирования в разрезе группы выборки по каждому из показателей.

Таблица 2.5 – Сравнение результатов работы методов прогнозирования для данных по общей заболеваемости (болезни, характеризующиеся повышенным кровяным давлением)

Группа выборки	Модель	RMSE	MAE	R2
Группа 1	ARIMA	4507.45	3429.32	-746.63
	HL	2750.56	2588.56	-277.4
	HW	26387.47	26316.0	-25621.27
	DT	1613.73	1494.75	-94.83
	RF	1613.73	1494.75	-94.83
	MLP	1313.71	1244.56	-62.51
Группа 2	ARIMA	1036.1	833.06	0.58
	HL	1078.48	1059.06	0.54
	HW	1094.06	1088.83	0.53
	DT	2241.18	1849.65	-0.98
	RF	2013.73	1525.0	-0.6
	MLP	2170.81	1558.26	-0.86
Группа 3	ARIMA	4658.81	4312.92	-1.82
	HL	7017.97	6556.78	-5.4
	HW	5005.91	4596.73	-2.26
	DT	6354.74	6349.14	-4.25
	RF	1747.76	1417.01	0.6
	MLP	2742.92	2717.65	0.02
Воронеж	ARIMA	2483.62	2251.73	0.53
	HL	3251.19	2416.48	0.2
	HW	2241.26	2199.66	0.62
	DT	2371.87	2365.05	0.57
	RF	3657.49	3104.67	-0.01
	MLP	4521.99	3915.95	-0.55
Область	ARIMA	1474.95	1377.22	0.69
	HL	2896.55	2248.09	-0.2
	HW	1454.84	1350.35	0.7
	DT	1712.81	1704.05	0.58
	RF	2568.42	2305.13	0.06
	MLP	3103.91	2483.83	-0.38

Таблица 2.6 – Сравнение результатов работы методов прогнозирования для данных по числу лиц, состоящих под диспансерным наблюдением на конец года (болезни, характеризующиеся повышенным кровяным давлением)

Группа выборки	Модель	RMSE	MAE	R2
Группа 1	ARIMA	444.98	317.58	-1.04
	HL	2040.95	1931.67	-41.87
	HW	23137.62	23084.62	-5509.15
	DT	217.52	157.95	0.51
	RF	1047.76	895.66	-10.3
	MLP	1912.71	1867.76	-36.66
Группа 2	ARIMA	517.86	442.95	0.92
	HL	2376.23	2131.54	-0.59
	HW	2958.41	2679.67	-1.46
	DT	1565.46	1452.37	0.31
	RF	1212.41	1208.36	0.59
	MLP	1899.57	1681.33	-0.01
Группа 3	ARIMA	6049.41	5364.68	-2.77
	HL	3429.11	2930.33	-0.21
	HW	4344.29	3716.67	-0.94
	DT	4515.17	4477.06	-1.1
	RF	2710.09	2625.0	0.24
	MLP	3430.36	3128.63	-0.21
Воронеж	ARIMA	2808.52	2598.72	0.6
	HL	3946.28	3092.46	0.21
	HW	3241.64	2792.87	0.47
	DT	4780.48	3681.55	-0.15
	RF	5211.71	3958.56	-0.37
	MLP	6595.78	4820.06	-1.2
Область	ARIMA	6365.82	5595.23	-3.11
	HL	3368.43	2687.3	-0.15
	HW	1810.44	1555.21	0.67
	DT	2056.75	1712.7	0.57
	RF	2718.25	2447.4	0.25
	MLP	3702.79	2977.16	-0.39

На основе полученных результатов сформирован ансамбль предобученных прогностических моделей с наилучшей точностью, представленный в таблице 2.7.

Таблица 2.7 – Ансамбль предобученных прогностических моделей с наилучшей точностью (болезни, характеризующиеся повышенным кровяным давлением)

Показатель	Группа выборки	Модель
Общая заболеваемость (табл. 2.5)	Группа 1	MLP
	Группа 2	ARIMA
	Группа 3	RF
	Воронеж	HW
	Область	HW
Первичная заболеваемость (табл. В.1)	Группа 1	HL
	Группа 2	ARIMA
	Группа 3	HL
	Воронеж	HW
	Область	ARIMA
Число больных с впервые в жизни установленным диагнозом, взятых под диспансерное наблюдение (табл. В.2)	Группа 1	MLP
	Группа 2	ARIMA
	Группа 3	MLP
	Воронеж	ARIMA
	Область	HW
Число больных с впервые в жизни установленным диагнозом, выявленных при профосмотре (табл. В.3)	Группа 1	-
	Группа 2	ARIMA
	Группа 3	MLP
	Воронеж	MLP
	Область	HW
Число больных с впервые в жизни установленным диагнозом, выявленных при диспансеризации (табл. В.4)	Группа 1	DT
	Группа 2	HL
	Группа 3	HL
	Воронеж	ARIMA
	Область	ARIMA
Число лиц, состоящих под диспансерным наблюдением на конец года (табл. 2.6)	Группа 1	DT
	Группа 2	ARIMA
	Группа 3	RF
	Воронеж	ARIMA
	Область	HW

В рамках исследования наборов данных и моделей прогнозирования рассмотрены возможности использования данных по разным территориальным образованиям и разным показателям эффективности.

В общем виде решаемую задачу предиктивного моделирования можно представить как задачу прогнозирования временного ряда:

$$y_j(t) = f \left(\begin{array}{c} y_j(t-1), y_j(t-2), \dots, y_j(t-p+1); \\ \theta \end{array} \right) + \varepsilon(t), \quad (2.5)$$

$$t = \overline{1, T}, j = \overline{1, J},$$

где:

f – модель прогнозирования;

p – размер временного окна для прогнозирования;

T – общее количество лет в наборе данных;

t – целевой год моделирования;

ε – ошибка прогноза;

θ – параметры модели

Предлагается постановка задачи с использованием данных по показателям в других территориальных образованиях:

$$y_j(t) = f \left(\begin{array}{c} y_j(t-1), y_j(t-2), \dots, y_j(t-p+1); \\ y_j^d(t-1), y_j^d(t-2), \dots, y_j^d(t-p+1); \\ \theta \end{array} \right) + \varepsilon(t), \quad (2.6)$$

$$t = \overline{1, T}, j = \overline{1, J}, d = \overline{1, D},$$

где d – территориальное образование;

y_j^d – показания по нозологии в год моделирования в территориальном образовании d ;

D – общее количество учитываемых при моделировании территориальных образований.

Для учета других показателей, не являющихся целевым, в постановку задачи добавлена информация о всех показателях в данном территориальном образовании:

$$y_j(t) = f \left(\begin{array}{l} y_j(t-1), y_j(t-2), \dots, y_j(t-p+1); \\ y_{j+l}(t-1), y_{j+l}(t-2), \dots, y_l(t-p+1); \\ \theta \end{array} \right) + \varepsilon_{t+1}, \quad (2.7)$$

$$t = \overline{1, T}, j = \overline{1, J}, l = \overline{1, J-1},$$

где l – коэффициент для извлечения в прогнозную модель не целевых для прогнозирования показателей.

2.3. Прогностическое моделирование интегрального эффекта от использования дополнительного ресурсного обеспечения

Для проведения прогностического моделирования интегрального эффекта от использования дополнительного ресурсного обеспечения требуется решить задачу регрессии. Перед обучением регрессионной модели составлены обучающие наборы данных для нозологии «Болезни, характеризующиеся повышенным кровяным давлением» в соответствии со значениями показателя общей заболеваемости и дополнительных объемов диспансеризации на основе массивов (2.3) и преобразований (2.4).

Таким образом, было получено 2 обучающих набора данных, каждый из которых состоит из следующих атрибутов:

- year – год наблюдений;
- R – значения объемов диспансеризации или изменение значений объемов диспансеризации;
- у – изменения показателя общей заболеваемости или отношение изменения показателя общей заболеваемости к изменению объемов диспансеризации (целевое значение регрессии);

– reg – численные представления территориального образования (1 – группа 1, 2 – группа 2, 3 – группа 3; 4 – ГО г. Воронеж, 5 – Область).

Обучающие наборы данных на основе данных многолетнего статистического учета в соответствии с массивами (2.3) содержат 45 строк, а на основе прогностических оценок, полученных из прогнозов ансамблей прогностических моделей (таблица 2.7), показателей эффективности и дополнительного ресурсного обеспечения в соответствии с преобразованиями (2.4) – 65 строк.

Обучение зависимостей (2.3) и (2.4) осуществлялось с использованием алгоритмов регрессии машинного обучения: дерево решений (DTR), случайный лес (RFR) и нейронной сети многослойный перцептрон (MLPR).

В таблице 2.8 представлены сравнительные результаты по качеству прогностического моделирования на основе регрессионных моделей в разрезе нозологий.

Таблица 2.8 – Сравнение результатов работы моделей регрессии при обучении зависимостей для нозологии «Болезни, характеризующиеся повышенным кровяным давлением»

Обучаемая зависимость	Регрессор	Метрики качества		
		RMSE	MAE	R2
$F(\Delta R(t), \Delta y_j(t))$	DTR	1564.21	1032.36	0.7
	RFR	1505.23	1077.42	0.73
	MLPR	1057.18	784.19	0.87
$F\left(\frac{\Delta y_j(t)}{\Delta R(t)}\right)$	DTR	0.36	0.28	-0.03
	RFR	0.31	0.24	0.23
	MLPR	100.02	88.15	-80947.54

Результаты оценки качества прогностического моделирования интегрального эффекта позволяют сделать вывод, что обученная зависимость (2.3) показала наилучший результат по точности в сравнении с обучением зависимости на основе преобразований (2.4) с использованием модели регрессии нейронной сети многослойный перцептрон и предлагается при построении единой прогностической модели в рамках горизонта прогнозирования для дальнейшей

оптимизации по величине интегрального эффекта от использования дополнительного ресурса в целях принятия решения по распределению ресурсного обеспечения между временными периодами.

Выводы второй главы

1. Для повышения эффективности использования результатов предиктивного анализа ретроспективной информации при принятии управленческих решений по распределению ресурсного обеспечения в региональных отраслевых организационных системах целесообразно объединить в рамках алгоритмической схемы экспертное оценивание на основе визуализации временных рядов исходных данных и прогностическое моделирование на основе машинного обучения.

2. Поскольку прогностическое моделирование на основе машинного обучения позволяет использовать разные модели, что отражается в оценке точности прогнозирования для каждого детализированного варианта тестовой выборки, приемлемым является формирование ансамбля предобученных прогностических моделей. Процесс дообучения осуществляется на выборках, необходимых для использования результатов предиктивного анализа при принятии управленческих решений: показателей эффективности, объемов дополнительного ресурсного обеспечения, выявленных значений интегрального эффекта от влияния дополнительного ресурса.

3. При выборе для моделирования с использованием средств интеллектуализации принятия решений временной зависимости интегрального эффекта от дополнительного ресурсного обеспечения в рамках горизонта прогнозирования целесообразен сравнительный анализ по точности двух способов

вычисления величины интегрального эффекта: на основе данных многолетнего статистического учета; на основе прогностических оценок показателей эффективности и дополнительного ресурсного обеспечения.

ГЛАВА 3. РАЗРАБОТКА ПРОЦЕДУР ИНТЕГРАЦИИ РЕЗУЛЬТАТОВ ПРЕДИКТИВНОГО АНАЛИЗА РЕТРОСПЕКТИВНОЙ ИНФОРМАЦИИ В ПРОЦЕСС ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ НА ОСНОВЕ ОПТИМИЗАЦИИ

В структурной модели интеллектуализации управления (см. п.1.3) первый уровень схемы принятия решений направлен на получение результатов предиктивного анализа в форме прогностических моделей временных рядов ретроспективной информации. На втором уровне осуществляется принятие управленческих решений путем интеграции в оптимизационное моделирование результатов первого уровня.

Первая процедура двухуровневой схемы принятия решений связана с принятием управленческого решения по распределению дополнительного ресурсного обеспечения между группами населения региональной отраслевой организационной системы. Основа этой процедуры базируется на численном методе рандомизированного поиска при получении решения задачи многоальтернативной оптимизации. Вторая процедура также является распределительной. Необходимо для каждой группы населения распределить ресурс между территориальными образованиями, что также соответствует модели многоальтернативной оптимизации. При этом следует учитывать, что каждый цикл принятия управленческих решений выполняется для фиксированного временного периода.

Для моделей многоальтернативной оптимизации требуется преобразование прогнозных оценок предиктивного анализа в экспертные оценки, позволяющие с помощью лингвистических переменных идентифицировать параметры целевой функции оптимизационных задач.

Решение задач оптимизации разделено для каждого дискретного временного периода $t = \overline{T_1 + 1, T_1 + T}$ с ориентацией на экспертные оценки потребности по результатам предиктивного анализа, но сами результаты не всегда приводят к наилучшему варианту распределения дополнительного ресурса. Поэтому целесообразно скорректировать полученное на основе прогностической модели значение интегрального эффекта, являющегося результатом предиктивного анализа. В этом случае критерий оптимизации является функцией временных периодов и приводит к многомерной задаче оптимизации. Алгоритмизация поиска оптимального варианта достигается за счет последовательного решения на множестве одномерных задач.

3.1. Управление распределением ресурсного обеспечения между группами населения региональной отраслевой организационной системы

Необходимо рассмотреть процесс распределения дополнительного ресурсного обеспечения для временных периодов $t = \overline{T_1 + 1, T_1 + T}$ между $n = \overline{1, N}$ группами населения региона и управления ими на основе моделей оптимизации с интеграцией результатов предиктивного анализа ретроспективной информации. Следует отметить, что при распределении дополнительного ресурсного обеспечения $\Delta R(t)$ между $n = \overline{1, N}$ группами населения возникает ситуация, когда эксперт в состоянии предсказать те группы, для которых распределение

$$\Delta R_n(t - 1), n = \overline{1, N}, \sum_{n=1}^N \Delta R_n(t - 1) = \Delta R(t - 1)$$

дало, по его мнению, недостаточный эффект по улучшению значений показателей эффективности. Адекватными оптимизируемыми переменными, характеризующими указанную неоднородность влияния $\Delta R_n(t-1)$, являются следующие альтернативные переменные

$$x_n = \begin{cases} 1, & \text{если из объема дополнительного ресурса } \Delta R(t) \text{ целесообразно} \\ & \text{выделить ресурс } \Delta R_n(t) \text{ для } n-\text{й группы населения,} \\ 0, & \text{в противном случае, } n = \overline{1, N} \end{cases} \quad (3.1)$$

Зависимость величины интегрального эффекта от оптимизируемых переменных (3.1) представляется в виде линейной функции, для которой коэффициенты $\alpha_n, n = \overline{1, N}$ получены на основе экспертных оценок приоритетности выделения ресурса для n -й группы населения:

$$\sum_{n=1}^N \alpha_n x_n \rightarrow \max \quad (3.2)$$

Для формирования зависимости граничного требования от оптимизируемых переменных (3.1) в линейной форме использованы экспертные оценки потребности в ресурсном обеспечении для n -й группы населения

$$\Delta \hat{R}_n(t), n = \overline{1, N}.$$

В результате получена следующая формализация граничного требования

$$\sum_{n=1}^N \Delta \hat{R}_n(t) x_n \leq \Delta R(t) \quad (3.3)$$

Объединение требования альтернативности распределения ресурса (3.1) и линейных зависимостей (3.2), (3.3) приводит к следующей модели многоальтернативной оптимизации [78]

$$\begin{aligned} & \sum_{n=1}^N \alpha_n x_n \rightarrow \max, \\ & \sum_{n=1}^N \Delta \hat{R}_n(t) x_n \leq \Delta R(t), t = \overline{T_1 + 1, T_1 + T}, \\ & x_n = \begin{cases} 1, \\ 0, & \text{в противном случае, } n = \overline{1, N}. \end{cases} \end{aligned} \quad (3.4)$$

На основе решения (3.4) $x_n^*, n = \overline{1, N}$ имеет следующее управляемческое решение

$$\Delta R_n(t) = \begin{cases} \Delta \hat{R}_n(t), t = \overline{T_1 + 1, T_1 + T}, \text{ если } x_n^* = 1 \\ 0, \text{ в противном случае, } n = \overline{1, N} \end{cases} \quad (3.5)$$

Предложена процедура интеграции результатов предиктивного анализа в вычислении коэффициентов $\alpha_n, n = \overline{1, N}$. Она основана на визуализации полученных результатов $\Delta R_n(t), y_{jn}(t), j = \overline{1, J}, n = \overline{1, N}, t = \overline{T_1 + 1, T_1 + T}$, предъявлении их эксперту и экспертного оценивания термов лингвистических переменных. Предлагается обосновать использование следующей лингвистической переменной

$$\langle \text{допустимо изменить } \Delta \hat{R}_n(t) \text{ с учетом} \\ \text{прогностической оценки } \hat{y}_j(t) \rangle \quad (3.6)$$

и соответствующей функции принадлежности.

Поскольку важно обеспечить последовательное уменьшение значений $y_j(t)$ в качестве функции принадлежности приемлемо использовать гауссовскую либо сигмоидальную функцию принадлежности.

Важным достоинством этих функций принадлежности является то, что их использование обеспечивает получение непрерывно дифференцируемых гиперповерхностей отклика нечеткой модели оценивания результатов предиктивного анализа при выборе экспертом градаций лингвистической переменной (3.6) следующих термов:

$$\vartheta_1 = (\text{увеличить, уменьшить, без результата});$$

$$\vartheta_2 = (\text{сильно, существенно, несколько, немного, мало}).$$

Однако гауссовые функции, будучи симметричными, подходят для представления внутренних нечетких множеств. В случае нечетких оценок приоритетности в выборе групп населения, на основе которых определяются коэффициенты оптимизационной модели (3.4) $\alpha_n, n = \overline{1, N}$ требуется предоставлять не только внутренние, но и крайние, т.е. слева и справа, множества. Такую возможность дает использование левой и правой сигмоидальных функций, представленных на рисунке 3.1.

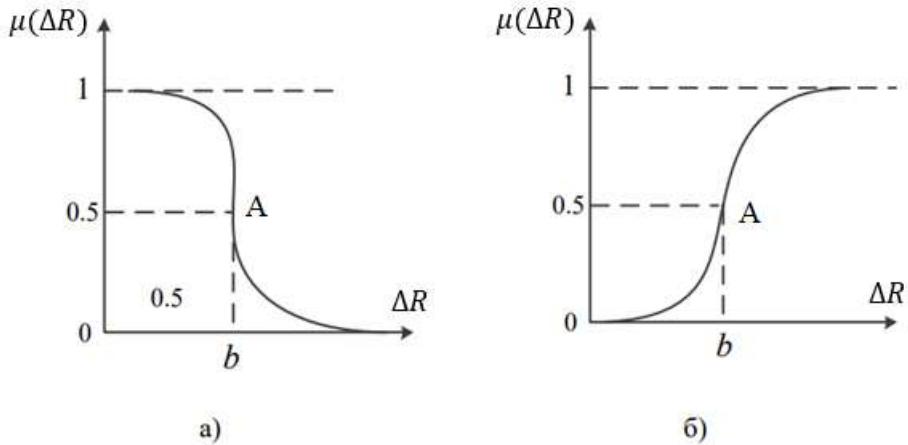


Рисунок 3.1 – Представление левой (а) и правой (б) сигмоидальных функций

Традиционно сигмоидальную функцию принадлежности $\mu(\Delta R)$ задают следующим выражением в зависимости от двух параметров (a, b)

$$\mu(\Delta R) = \frac{1}{1 + \exp[-a(\Delta R - b)]}, \quad (3.7)$$

где a – коэффициент наклона функции в точке перегиба A ;

$a > 0$ – правая сигмоидальная функция;

$a < 0$ – левая сигмоидальная функция;

b – координата точки A , принадлежащей нечеткому множеству со степенью 0,5.

Для более эффективного учета крайних множеств предлагается использовать семейство трехпараметрических (a, b, c) сигмоидальных функций, построенных на основе экспоненциальной зависимости от параметров (a, b) в знаменателе (3.7), степенной зависимости $(\Delta R^{\vartheta} n_j - c)$, где $\Delta R^{\vartheta} n_j$ – относительное значение дополнительного ресурсного обеспечения для n -й группы населения

$$\Delta R^{\vartheta} n_j = \frac{\Delta R^{\vartheta} n_j}{\Delta \hat{R} n_j},$$

где $\Delta \hat{R} n_j$ – экспертная оценка потребности в ресурсном обеспечении;

$\Delta R^{\vartheta} n_j$ – объем дополнительного ресурсного обеспечения, предлагаемый экспертом как изменения ΔR_n на основе предиктивного анализа j -му показателю $\Delta y_{nj}(t), j = \overline{1, J}, n = \overline{1, N}, t = \overline{T_1 + 1, T_1 + T}$.

Функции принадлежности, входящие в семейство, различаются зависимостью от параметров (a, b, c) с учетом значений термов лингвистической переменной ϑ_1 , а сами параметры устанавливаются таблицей с учетом сочетаний значений термов ϑ_1 и ϑ_2 :

$$\vartheta_1 = \langle \text{увеличить} \rangle$$

$$\mu'_{nj} = \begin{cases} 1, & \text{если } \Delta R^\circ_{nj} \geq b, \\ \frac{1}{1 + [a(\Delta R^\circ_{nj} - c)]^b}, & \text{если } \Delta R^\circ_{nj} < b \end{cases}$$

$$\vartheta_1 = \langle \text{уменьшить} \rangle$$

$$\mu''_{nj} = \begin{cases} 0, & \text{если } \Delta R^\circ_{nj} \leq c, \\ 1 - \frac{1}{1 + [a(\Delta R^\circ_{nj} - c)]^b}, & \text{если } \Delta R^\circ_{nj} > c \end{cases}$$

$$\vartheta_1 = \langle \text{безразлично} \rangle$$

$$\mu_{nj} = \begin{cases} \mu'_{nj}, & \text{если } \Delta R^\circ_{nj} < b \\ 1, & \text{если } b \leq \Delta R^\circ_{nj} \leq c \\ \mu''_{nj}, & \text{если } \Delta R^\circ_{nj} > c \end{cases}$$

В результате определено фиксированное значение функции принадлежности

$$\mu_{nj}(t), n = \overline{1, N}, j = \overline{1, J}, t = \overline{T_1 + 1, T_1 + T}. \quad (3.8)$$

Коэффициент α_n вычисляется на основе усреднения значений (3.8) на множестве $j = \overline{1, J}$

$$\alpha_n(t) = \frac{\sum_{j=1}^J \mu_{nj}(t)}{J}, n = \overline{1, N}, t = \overline{T_1 + 1, T_1 + T}.$$

Структурная схема процедуры формирования оптимизационной модели с использованием результатов предиктивного анализа приведена на рисунке 3.2.



Рисунок 3.2 – Структурная схема процедуры формирования оптимизационной модели с использованием результатов предиктивного анализа

Для получения решения $x^*, n = \overline{1, N}$ оптимизационной задачи целесообразным является применение рандомизированного поиска на основе замены переменных (3.1) на случайные величины, значения которых зависят от следующих вероятностных характеристик

$$P(\tilde{x}_n = 1) = p_{x_n}, P(\tilde{x}_n = 0) = q_{x_n}, p_{x_n} + q_{x_n} = 1; \quad (3.9)$$

$$P(\tilde{n} = n) = p_n, n = \overline{1, N}, \sum_{n=1}^N p_n = 1. \quad (3.10)$$

где P – обозначения оценки вероятности события;

$\tilde{x}_n = 1$ – случайное событие $x_n = 1$,

$\tilde{x}_n = 0$ – случайное событие $x_n = 0$,

$$\tilde{x}_n = \begin{cases} 1, & \text{если } \tilde{\varepsilon} \leq p_{x_n} \\ 0, & \text{в противном случае} \end{cases},$$

$$\tilde{n} = \begin{cases} 1, & \text{если } \tilde{\varepsilon} \leq p_1 \\ 2, & \text{если } \tilde{\varepsilon} \leq p_1 + p_2, \\ \dots \\ n, & \text{если } \tilde{\varepsilon} \leq 1 \end{cases}$$

где $\tilde{\varepsilon}$ – последовательность псевдослучайных чисел, равномерно распределенных на интервале $(0,1)$.

На первом шаге итерационного процесса поиска $x^*, n = \overline{1, N}$ при $k = 1$ принимают равномерные распределения (3.9), (3.10). Если у экспертов есть предпочтения по некоторым группам населения, то для них $p_{x_n} > 0.5$, $p_n = \frac{1}{N}$, то есть распределение (3.9) и распределение (3.10) становятся неравномерными.

На k -й итерации получены вероятностные характеристики $P^k_{x_n}, P^k_n, n = \overline{1, N}$, которые изменяются на $(k + 1)$ -й итерации в зависимости от случайных реализаций вариации оптимизируемой функции объединяющей целевую функцию и ограничение в задаче (3.4):

$$\varphi_{n_1} = \varphi(\tilde{x}_n(n \neq n_1), x_{n_1} = 1) - \varphi(\tilde{x}_n(n \neq n_1), x_{n_1} = 0),$$

где

$$\varphi(x_n) = \sum_{n=1}^N \alpha_n x_n + y(\Delta R - \sum_{n=1}^N \Delta \hat{R} x_n),$$

$y^1 > 0$ – устанавливается экспертом при $k = 1$,

$$y^{k+1} = \max \left(0, y^k - \gamma \left(\Delta R - \sum_{n=1}^N \Delta \hat{R} x_n^k \right) \right),$$

$\gamma > 0$ – величина шага, устанавливается экспертом при $k = 1$.

Правило останова итерационного процесса поиска $k = K$ – заданное число итераций позволяет выделить $l = \overline{1, L}$ доминирующих вариантов наборов альтернативных переменных

$$x_{nl}, n = \overline{1, N}, l = \overline{1, L}. \quad (3.11)$$

Для выбора окончательного управлеченческого решения на множестве вариантов (3.11) используется экспертная информация. В отличие от приведенных в [79] процедур предлагается ориентировать эксперта на результаты предиктивного анализа в виде графического представления, для каждого l -го варианта функций

$$\begin{aligned} \Psi_1(t) &= \sum_{n=1}^N \Delta R_n(t) x_{nl}(t), l = \overline{1, L}, t = \overline{T_1 + 1, T_1 + T}; \\ \Psi_2(t) &= \sum_{n=1}^N \Delta \hat{R}_n(t) x_{nl}(t), l = \overline{1, L}, t = \overline{T_1 + 1, T_1 + T}; \end{aligned}$$

и их визуального сравнения. Если на основе визуального сравнения эксперт не приходит к однозначному решению l^* , то выбор осуществляется по условию

$$\min_{l = \overline{1, L}} \frac{\sum_{t=T_1+1}^{T_1+T} |\Psi_1(t) - \Psi_2(t)|}{T}$$

с определением l^* .

Таким образом выбирается вариант $x_{nl^*}(t)$, которому соответствует управлеченческое решение

$$\Delta R_n(t) = \begin{cases} \Delta \hat{R}_{n_1}(t), & \text{если } x_{n_1 l^*}(t) = 1, t = \overline{T_1 + 1, T_1 + T}, n = \overline{1, N} \\ 0, & \text{в противном случае} \end{cases} \quad (3.12)$$

Структурная схема процедуры принятия управленческого решения при распределении ресурсного обеспечения между группами населения региона приведена на рисунке 3.3.

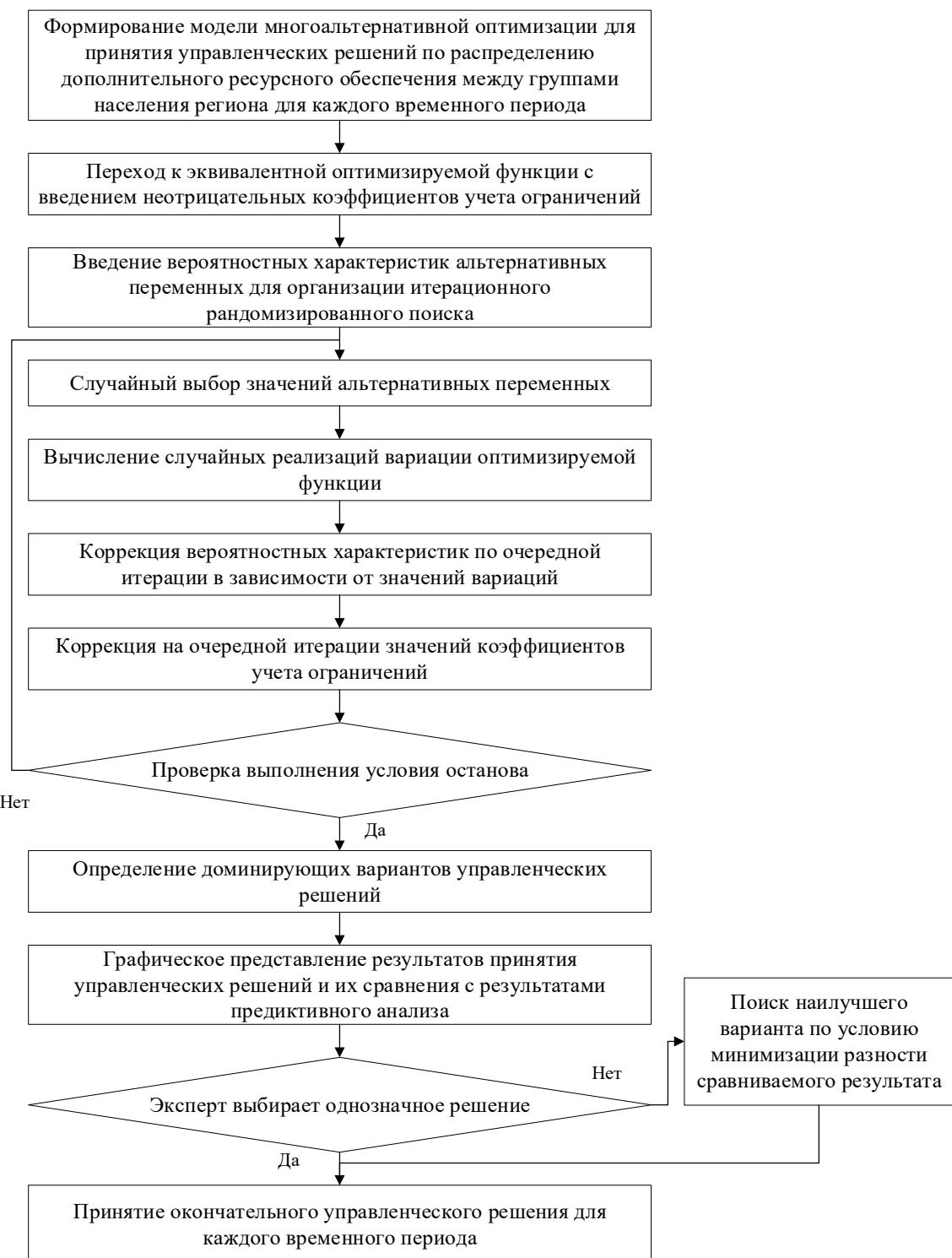


Рисунок 3.3 – Структурная схема процедуры принятия управленческого решения при распределении дополнительного ресурсного обеспечения между группами населения региона

3.2. Управление распределением ресурсного обеспечения между территориальными образованиями региональной отраслевой организационной системы

Вторая задача распределения, исходя из решения (3.12), состоит в распределении тех $\Delta R_n(t)$, для которых оптимальное значение

$$x_{n_1}^* = 1, n_1 = \overline{1, N} \in \overline{1, N},$$

между $d = \overline{1, D}$ территориальными образованиями. Здесь, как и в первой задаче, возникает возможность экспертного ранжирования территорий по значимости влияния на достижение требований управляющего центра. Тогда в качестве оптимизируемых вновь приемлемо ввести альтернативные переменные:

$$x_{nd} = \begin{cases} 1, & \text{если из объема дополнительного ресурса } \Delta R_n(t) \\ & \text{целесообразно выделить ресурс } \Delta R_{nd}(t) \\ & \text{для } d - \text{го территориального образования} \\ 0, & \text{в противном случае, } n = \overline{1, N}, d = \overline{1, D} \end{cases} \quad (3.13)$$

Далее определяются коэффициенты линейных зависимостей экстремального и граничных требований:

– коэффициенты влияния ресурса $\Delta \hat{R}_n(t)$ на достижение требований управляющего центра

$$\beta_{nd}, n = \overline{1, N}, d = \overline{1, D};$$

– экспертные оценки потребности в ресурсном обеспечении для d -го территориального образования

$$\Delta R_{nd}(t), n = \overline{1, N}, d = \overline{1, D}.$$

Линейные зависимости имеют вид аналогичный (3.2), (3.3), а оптимизационная модель на их основе – (3.4):

$$\sum_{d=1}^D \sum_{n=1}^N \beta_{nd} x_{nd} \rightarrow \max,$$

$$\sum_{d=1}^D \Delta \hat{R}_{n_1 d}(t) x_{nd} \leq \Delta R_{n_1}(t), n_1 = \overline{1, N_1}, t = \overline{T_1 + 1, T_1 + T} \quad (3.14)$$

$$x_{nd} = \begin{cases} 1, & n = \overline{1, N}, d = \overline{1, D} \\ 0, & \text{в противном случае} \end{cases}.$$

Решая задачу многоальтернативной оптимизации (3.14), получены оптимальные значения переменных $x_{nd}^*, n = \overline{1, N}, d = \overline{1, D}$ и следующее управленическое решение

$$\Delta R_{nd}(t) = \begin{cases} \Delta R_{n_1}(t), & t = \overline{T_1 + 1, T_1 + T}, \text{ если } x_{n_1 d}^* = 1 \\ 0, & \text{в противном случае} \end{cases}, \quad (3.15)$$

$$n_1 = \overline{1, N_1}, n = \overline{1, N}, n_1 = \overline{1, N_1} \in \overline{1, N}, d = \overline{1, D}.$$

В соответствии с управленическим решением (3.15) из объема ресурсного обеспечения ΔR_{n_1} d_1 -му территориальному образованию выделяется ресурс $\Delta R_{n_1 d_1}(t)$, где $d_1 = \overline{1, D_1} \in \overline{1, D}$ и соответствует $x_{n_1 d_1}^* = 1$.

Экспертные оценки формируются на основе визуализации результатов предиктивного анализа временных графиков, полученных путем прогностического моделирования

$$\begin{aligned} \Delta y_{jn_1 d}(t), & j = \overline{1, J}, n_1(t) = \overline{1, N_1(t)}, d = \overline{1, D}, t = \overline{T_1 + 1, T_1 + T}, \\ \Delta R_{n_1 d}(t), & n_1(t) = \overline{1, N_1(t)}, d = \overline{1, D}, t = \overline{T_1 + 1, T_1 + T}. \end{aligned}$$

Коэффициенты $\beta_{n_1 d}$ определяются так же, как в первой задаче управления с использованием нечетких множеств соответствующих лингвистической переменной

$$\begin{aligned} <\text{допустимо изменить } \Delta \hat{R}_{n_1 d}(t) \text{ с учетом прогностической} \\ & \text{оценки } \Delta y_{jn_1 d}(t)>. \end{aligned} \quad (3.16)$$

С использованием градаций термов (3.16), которые указывает эксперт, и семейства сигмоидальных функций принадлежности определяются ее значения

$$\mu_{n_1 d_j}(t), n_1(t) = \overline{1, N_1(t)}, d = \overline{1, D}, t = \overline{T_1 + 1, T_1 + T}. \quad (3.17)$$

Значения $\mu_{n_1 d_j}(t)$ определяются для относительных значений дополнительного ресурсного обеспечения

$$\Delta R^{\circ}_{n_1 d_j} = \frac{\Delta R^{\vartheta}_{n_1 d_j}}{\Delta \hat{R}_{n_1 d}},$$

где $\Delta \hat{R}_{n_1 d}$ – экспертная оценка потребности в дополнительном ресурсном обеспечении;

$\Delta R^{\vartheta}_{n_1 d_j}$ – объем дополнительного ресурсного обеспечения, предлагаемый экспертом как изменение $\Delta \hat{R}_{n_1 d}$ на основе предиктивного анализа по j -му показателю $\Delta y_{n_1 d_j}(t), j = \overline{1, J}, n_1(t) = \overline{1, N_1(t)}, d = \overline{1, D}, t = \overline{T_1 + 1, T_1 + T}$.

В выражениях для функций принадлежности основанием для степенной функции является значение $(\Delta R^{\circ}_{n_1 d_j} - c)$.

Коэффициенты $\beta_{n_1 d}(t)$ вычисляются на основе усредненных значений (3.17) на множестве $j = \overline{1, J}$

$$\beta_{n_1 d}(t) = \frac{\sum_{j=1}^J \mu_{n_1 d_j}(t)}{J}, n_1 = \overline{1, N_1}, d = \overline{1, D}, t = \overline{T_1 + 1, T_1 + T}.$$

Процедуры рандомизированного поиска управлеченческого решения (3.15) на основе оптимизационной модели (3.14) аналогична процедуре, использованной при решении первой задачи управления, что позволяет при $k = K$ выделить $l = \overline{1, L}$ доминирующих вариантов наборов альтернативных переменных

$$x_{n_1 d l}, n_1 = \overline{1, N_1}, d = \overline{1, D}, l = \overline{1, L}. \quad (3.18)$$

На следующем этапе экспертного выбора используются результаты предиктивного анализа в виде графической визуализации для каждого l -го варианта следующих функций:

$$\begin{aligned} \varphi_1(t) &= \sum_{d=1}^D \Delta R_{n_1 d}(t) x_{n_1 d l}(t), n_1(t) = \overline{1, N_1(t)}, t = \overline{T_1 + 1, T_1 + T}, l \\ &= \overline{1, L}, d = \overline{1, D}. \end{aligned} \quad (3.19)$$

$$\begin{aligned} \varphi_2(t) &= \sum_{d=1}^D \Delta \hat{R}_{n_1 d}(t) x_{n_1 d l}(t), n_1(t) = \overline{1, N_1(t)}, t = \overline{T_1 + 1, T_1 + T}, l \\ &= \overline{1, L}, d = \overline{1, D}. \end{aligned} \quad (3.20)$$

Этап окончательного выбора аналогичен первой задаче управления, но с использованием функций (3.19), (3.20) и определения $l^* \in \overline{1, L}$.

То есть выбирается вариант $x_{n_1 d_1 l^*}(t)$, которому соответствует управлеченческое решение

$$\Delta R_n(t) = \begin{cases} \Delta \hat{R}_{n_1 d_1}(t), & \text{если } x_{n_1 d_1 l^*}(t) = 1, t = \overline{T_1 + 1, T_1 + T} \\ 0, & \text{в противном случае} \end{cases}. \quad (3.21)$$

Структурная схема процедуры формирования оптимизационной модели и принятия управлеченческого решения при распределении ресурсного обеспечения между территориальными образованиями региональной отраслевой организационной системы приведена на рисунке 3.4.



Рисунок 3.4 – Структурная схема процедуры формирования оптимизационной модели и принятия управлеченческого решения при распределении ресурсного обеспечения между территориальными образованиями региональной отраслевой организационной системы

3.3. Управление распределением ресурсного обеспечения между временными периодами в рамках горизонта прогнозирования

Управленческие решения (3.12), (3.21) базируются на экспертных оценках потребности в ресурсном обеспечении и позволяют определить подмножество групп населения $n_1 = \overline{1, N_1} \in \overline{1, N}$ и соответствующее подмножество территориальных образований $d_1 = \overline{1, D_1} \in \overline{1, D}$, для которых в соответствии с экспертными оценками $\alpha_n, n = \overline{1, N}$ и $\beta_{nd}, d = \overline{1, D}$ запланированный ресурс позволит улучшить значения показателей эффективности $y_j, j = \overline{1, J}$. Однако ориентация только на мнения экспертов не всегда приводит к оптимальному распределению интегрированного объема ресурсного обеспечения для каждого временного периода $t = \overline{T_1 + 1, T_1 + T}$.

Предлагается осуществить коррекцию распределения $\Delta R_{nd}(t)$ по временным периодам $t = \overline{T_1 + 1, T_1 + T}$ на основании формализованной постановки оптимизационной задачи на множестве дискретных значений

$$\Delta R_{n_1 d_1}^{\text{ck}}(t) = \{\Delta \overline{R_{n_1 d_1}}, 2\Delta \overline{R_{n_1 d_1}}, \dots, t\Delta \overline{R_{n_1 d_1}}, \dots, T\Delta \overline{R_{n_1 d_1}}\}, \quad (3.22)$$

где $\Delta \overline{R_{n_1 d_1}}$ среднее значение ресурсного обеспечения за период $t = \overline{T_1 + 1, T_1 + T}$, выбранного на основе управленческих решений (3.12, 3.21)

$$\Delta \overline{R_{n_1 d_1}} = \frac{\sum_{t=T_1+1}^{T_1+T} \Delta R_{n_1 d_1}(t)}{T}.$$

Экстремальное требование заключается в максимизации функции интегрального эффекта, а граничное – интегральным объемом дополнительного ресурсного обеспечения, полученного на основе управленческих решений (3.12), (3.21)

$$\Delta R_{nd} = \sum_{t=T_1+1}^{T_1+T} \Delta R_{n_1 d_1}. \quad (3.23)$$

С учетом функций интегрального эффекта (3.22), (3.23) оптимизационная модель имеет следующий вид [79]:

$$\begin{aligned}
 & F(\Delta R_{n_1 d_1}^{\text{ck}}(t); \Delta y_j(t), j = \overline{1, J}; t) \rightarrow \max, \\
 & \sum_{t=T_1+1}^{T_1+T} \Delta R_{n_1 d_1}^{\text{ck}}(t) = \Delta R_{n_1 d_1}, \\
 & \Delta R_{n_1 d_1}^{\text{ck}}(t) = \{\overline{\Delta R_{n_1 d_1}}, 2\overline{\Delta R_{n_1 d_1}}, \dots, t\overline{\Delta R_{n_1 d_1}}, \dots, T\overline{\Delta R_{n_1 d_1}}\}.
 \end{aligned} \tag{3.24}$$

В оптимизационной модели (3.24) результаты предиктивной аналитики присутствуют при машинном обучении зависимости интегрального эффекта от дополнительного ресурса и временных периодов. В этом случае модель регрессии помимо факторов ΔR_{nd} и $\Delta y_j(t)$ включает в себя время, что позволяет ее рассматривать как в качестве функции, которую необходимо максимизировать на множестве временных значений оптимизируемых переменных, так и в качестве прогностической функции изменения интегрального эффекта в зависимости от значений дополнительного ресурсного обеспечения. Такой характер данной функциональной модели дает возможность интегрировать результаты предиктивной аналитики в процесс принятия управленческих решений $\Delta R_{n_1 d_1}^{\text{ck}*}(t)$.

С этой целью переходят к последовательному решению одномерных задач оптимизации функций, которые вычисляются при значениях $\Delta R_{n_1 d_1}^{\text{ck}}(t)$ для определенного временного периода с использованием обученной на данных многолетней статистики зависимости интегрального эффекта от дополнительного ресурса и временных периодов. Указанные значения $\Delta R_{n_1 d_1}^{\text{ck}}(t)$ выбираются путем организации двухуровневого перебора для поиска максимального из возможных значения одномерной функции.

На первом уровне осуществляется перебор значений $\Delta R_{n_1 d_1}^{\text{ck}}(t)$ на интервале

$$\overline{\Delta R_{n_1 d_1}} \leq \Delta R_{n_1 d_1}^{\text{ck}}(t) \leq Z_{t'}, \tag{3.25}$$

где $Z_{t'}, t' = \overline{2, T} + 1$ – варьируемое значение верхней границы интервала (3.25). На втором уровне проводится перебор значений $Z_{t'}$ на множестве дискретных значений (3.22).

Двухуровневый перебор позволяет осуществить поиск $\Delta R_{n_1 d_1}^{\text{CK*}}(t), t = \overline{T_1 + 1}, \overline{T_1 + T}$ на основе следующей совокупности задач максимизации одномерных функций

$$\begin{aligned} & \varphi_{t'}[Z_{t'}, \Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + t'), \varphi_{t'-1}], \\ & \max \varphi_{t'}[\Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + t'), t = T_1 + t'], \\ & \Delta \overline{R_{n_1 d_1}} \leq \Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + t') \leq Z_{t'}, \end{aligned} \quad (3.26)$$

где

$$\begin{aligned} \varphi_1(Z_2) &= F[\Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + 1), t = T_1 + 1], \\ \varphi_2(Z_3) &= F[\Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + 2), t = T_1 + 2] + \varphi_1[Z_3 - \Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + 2)], \\ & \vdots \\ \varphi_T(Z_{T+1}) &= F[\Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + T), t = T_1 + T] + \varphi_{T-1}[Z_{T+1} - \\ & \quad \Delta R_{n_1 d_1}^{\text{CK}}(t = T_1 + T)]. \end{aligned}$$

Получение решения (3.26) позволяет определить окончательное управлеченческое решение

$$\Delta R_{n_1 d_1}^{\text{CK*}}(t), n_1 = \overline{1, N_1}, d_1 = \overline{1, D_1}, t = \overline{T_1 + 1, T_1 + T}.$$

Структурная схема процедуры формирования оптимизационной модели и принятия скорректированного управлеченческого решения по распределению дополнительного ресурсного обеспечения между временными периодами на основе предиктивного анализа приведена на рисунке 3.5.



Рисунок 3.5 – Структурная схема процедуры формирования оптимизационной модели и принятия скорректированного управлеченческого решения по распределению дополнительного ресурсного обеспечения между временными периодами на основе предиктивного анализа

Выводы третьей главы

1. Процедуру принятия решений при управлении распределением ресурсного обеспечения между группами населения региональной отраслевой организационной системы целесообразно сформировать из двух частей:

формирование модели многоальтернативной оптимизации и итерационного поиска. При этом для каждой части повышения эффективности процесса выбора способствует интеграция результатов предиктивного анализа ретроспективной информации. Для формализации экстремального требования указанные результаты преобразуются в значения функций принадлежности, полученные на основе экспертных оценок термов лингвистических переменных. Поскольку итерационный поиск завершается формированием множества доминирующих вариантов, окончательное управлеченческое решение определяется экспертом на основе прогностической модели интегрального эффекта. Процедура применяется для каждого временного периода.

2. Процедуру управления распределением ресурсного обеспечения между территориальными образованиями региональной отраслевой организационной системы приемлемо построить по аналогии с процедурой принятия решений по распределению между группами населения. Отличие заключается в том, что задача оптимизации решается только для тех групп, которые вошли в предыдущее решение и визуализируются для экспертного оценивания результатов предиктивного анализа, детализированных по территориальным образованиям.

3. Процедура управления распределением ресурсного обеспечения между временными периодами в рамках горизонта прогнозирования позволяет скорректировать детализованные решения для групп населения и территориальных образований путем исключения экспертного оценивания и использования единой прогностической модели интегрального эффекта в рамках горизонта прогнозирования. Оптимизация осуществляется за счет перехода от многомерной оптимационной модели к семейству одномерных и применения двухуровневого перебора значений дополнительного ресурсного обеспечения и интервала его изменения на каждом временном периоде.

ГЛАВА 4. ПРИМЕНЕНИЕ РАЗРАБОТАННЫХ СРЕДСТВ ИНТЕЛЛЕКТУАЛИЗАЦИИ ПРИНЯТИЯ РЕШЕНИЙ В ПРАКТИКЕ УПРАВЛЕНИЯ РЕГИОНАЛЬНОЙ СИСТЕМОЙ ЗДРАВООХРАНЕНИЯ

Применение разработанных средств интеллектуализации принятия решений в практике управления в региональных отраслевых организационных системах основывается на прогнозной реализации предложенных процедур предиктивного анализа, экспертного оценивания и оптимизации. Структура программного обеспечения определяет взаимодействие информационных ресурсов, оригинальных программных средств, на которые получены свидетельства о государственной регистрации, и стандартных программных продуктов машинного обучения и оптимизации. При этом данная структура должна обеспечивать реализацию структурной модели управления распределением ресурсного обеспечения с использованием двухуровневой схемы принятия решений.

Созданное программное обеспечение дает возможность оценить практическое применение интеллектуальной поддержки принятия решений при управлении процессом диспансеризации в региональной системе здравоохранения Воронежской области. Ретроспективная информация за 10 временных периодов (календарных лет) позволяет апробировать подходы к предиктивному анализу и оптимизации, предложенные в главах 2 и 3. Полный массив данных и рассмотренная в главе 2 тестовая выборка, направлены на построение прогностических моделей в рамках заданного горизонта прогнозирования путем дообучения предобученных моделей, объединенных в ансамбль по наилучшей точности прогнозирования.

Наличие модельной среды создает предпосылки для проведения вычислительного эксперимента по принятию управленческих решений о

распределении дополнительных объемов диспансеризации на основе интеграции результатов предиктивного анализа и оптимизации. В этом случае возникает возможность сравнительной оценки результатов управления с использованием с одной стороны, средств интеллектуализации, а с другой – традиционных экспертных решений управляющего центра по величине интегрального эффекта и времени, затрачиваемого на принятие решений.

4.1. Структура программного обеспечения

Комплекс программного обеспечения состоит из трех компонентов:

- разработанное на языке программирования Python программное обеспечение классификации данных (ПОКД), на которое получены два свидетельства о государственной регистрации в реестре Федеральной службы по интеллектуальной собственности;
- разработанные блоки программного кода для прогностического моделирования в среде разработки Jupyter-Notebook, которая позволяет видеть результат выполнения кода и его отдельных фрагментов в произвольном порядке;
- разработанные блоки таблиц для построения графиков в целях визуального моделирования и проведения оптимизационного моделирования в среде Microsoft Excel, которое представляет собой стандартное программное обеспечение для работы с электронными таблицами, позволяющее организовывать, анализировать и визуализировать данные.

Предварительным этапом перед использованием разработанного программного инструментария является формирование выборок многолетней

статистической информации на основе проведенного статистического учета в региональной организационной системе.

В разработанном ПОКД были использованы следующие компоненты:

1) высокоуровневый язык программирования (ЯП) общего назначения Python;

2) PyCharm – интегрированная среда разработки для ЯП Python.

Предоставляет средства для анализа кода, графический отладчик, инструмент для запуска юнит-тестов;

3) PyQt5 – библиотека, которая позволяет использовать фреймворк Qt GUI (графический интерфейс пользователя) в Python. Является одним из наиболее часто используемых модулей для создания GUI приложений в Python [90];

4) PyQt5 Designer – библиотека, необходимая для визуального программирования интерфейса программного обеспечения;

5) csv – модуль, в котором реализованы классы для чтения и записи табличных данных в формате .csv;

6) pandas – программная библиотека на ЯП Python для обработки и анализа данных;

7) scikit-learn – бесплатная библиотека машинного обучения для ЯП Python;

8) библиотека pickle, реализующая двоичные протоколы для сериализации и десериализации структуры объектов Python [91].

ПОКД предназначено для проведения бинарной и многоклассовой классификации объектов из различных прикладных областей путем построения и обучения классификационных моделей разного уровня сложности и оценки качества этих моделей.

Программа позволяет обучать классификаторы, выбирать их, настраивать параметры вручную, либо проводить обучение моделей с использованием метода поиска по сетке при подборе параметров. Загружать наборы данных для обучения и для прогноза. Сохранять обученные классификационные модели и использовать их для прогнозирования. Проводить оценку качества, просматривать и сохранять историю обучений. Задача классификации может решаться следующими

алгоритмами: k-ближайших соседей, логистическая регрессия, наивный Байес, дерево решений, многослойный перцептрон, случайный лес, бэггинг, адаптивный бустинг, CatBoost, градиентный бустинг и стекинг. На рисунке 4.1 представлен интерфейс разработанного программного обеспечения.

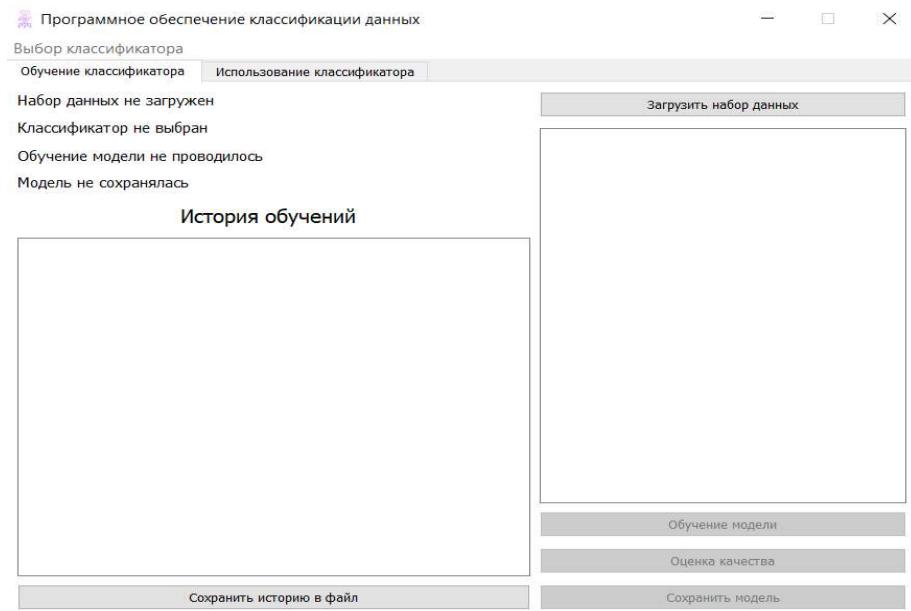


Рисунок 4.1 – Главная форма ПОКД

Результатом работы ПОКД является сформированный файл в формате CSV с распределенными по классам территориальных образований наборами признаков. Данный файл, вместе с исходными файлами многолетней статистической информации, загружается в среду разработки Jupyter-Notebook, в котором разработанные программные блоки обрабатывают расставленные классы объектов региональной организационной системы и распределяют данные статистической информации по классам, рассчитывая среднее арифметическое значений показателей эффективности и объемов ресурсного обеспечения в разрезе территориальных образований, по которым было осуществлено распределение классов, и сохраняются в виде новых файлов в формате CSV.

Следующим этапом является обработка данных многолетней статистической информации с учетом распределенных по классам территориальных образований путем загрузки сохраненного файла в формате CSV и проведения обучения

прогностических моделей: функциональных, нейросетевого, на основе деревьев решений и случайного леса с оценкой качества по трем критериям – R2, RMSE, MAE. На основе показателей качества формируется ансамбль предобученных моделей для каждого территориального образования региональной организационной системы. На основе сформированного ансамбля происходит дообучение моделей и формирование прогнозных значений на заданный горизонт планирования по моделям прогнозирования с наилучшим качеством по точности, которые сохраняются в файл в формате CSV для последующего использования в оптимизационном моделировании в стандартном программном продукте Microsoft Excel. Данная процедура выполняется отдельно для каждого показателя эффективности (или объемов ресурсного обеспечения) и группы населения.

На рисунке 4.2 приведена структурная схема взаимодействия разработанных программных продуктов с библиотеками стандартных программ машинного обучения и оптимизации.



Рисунок 4.2 – Структурная схема взаимодействия разработанных программных продуктов с библиотеками стандартных программ машинного обучения и оптимизации

4.2. Визуальное и прогностическое моделирование заболеваемости артериальной гипертензией лиц старших возрастных групп и их диспансеризации в системе здравоохранения Воронежской области

На примере региональной организационной системы здравоохранения (см. п.2.2) необходимо осуществить анализ динамики показателей заболеваемости по пяти нозологиям артериальной гипертензии для лиц старших возрастных групп и объемов их диспансеризации на основе ретроспективной информации, решив следующие задачи:

- формирование результатов анализа ретроспективной информации на основе визуального моделирования;
- формирование результатов анализа ретроспективной информации на основе прогностического моделирования.

Визуальный анализ основан на использовании эффективных механизмов наглядно-образной интуиции человека при сравнении временных графиков динамических процессов. С этой целью осуществляется визуализация временных рядов медико-статистической информации за 2013-2022 годы. Полученные графики позволяют сопоставить тенденции многолетней динамики:

- заболеваемости пятью нозологиями артериальной гипертензии лиц старшего и пожилого возраста в целом по региону и отдельно проживающих в областном центре и районах области, разделенных на группы по уровням заболеваемости;
- общей и первичной заболеваемости по пяти нозологиям артериальной гипертензии, характерным для взрослого населения;
- общей заболеваемости по пяти нозологиям артериальной гипертензии и числа больных, взятых под диспансерное наблюдение с выделением лиц старшего возраста, выявленных при диспансеризации и профосмотре.

Таким образом, сформированы следующие показатели:

- $y_1(t)$ – общая заболеваемость;
- $y_2(t)$ – первичная заболеваемость;
- $R_1(t)$ – число лиц, состоящих под диспансерным наблюдением на конец года;
- $R_2(t)$ – число больных с впервые в жизни установленным диагнозом, взятых под диспансерное наблюдение;
- $R_3(t)$ – число больных с впервые в жизни установленным диагнозом, выявленных при профосмотре;
- $R_4(t)$ – число больных с впервые в жизни установленным диагнозом, выявленных при диспансеризации.

Перечисленные показатели исследуются по следующим нозологиям:

- $n = 1$ – болезни, характеризующиеся повышенным кровяным давлением;
- $n = 2$ – эссенциальная гипертензия;
- $n = 3$ – гипертензивная болезнь сердца;
- $n = 4$ – гипертензивная болезнь с преимущественным поражением почек;
- $n = 5$ – гипертензивная болезнь с преимущественным поражением сердца и почек.

Для визуального анализа показателей заболеваемости артериальной гипертензией взрослого населения региона на примере двух нозологий осуществлено графическое представление временных рядов, приведенных на рисунках 4.3-4.14.

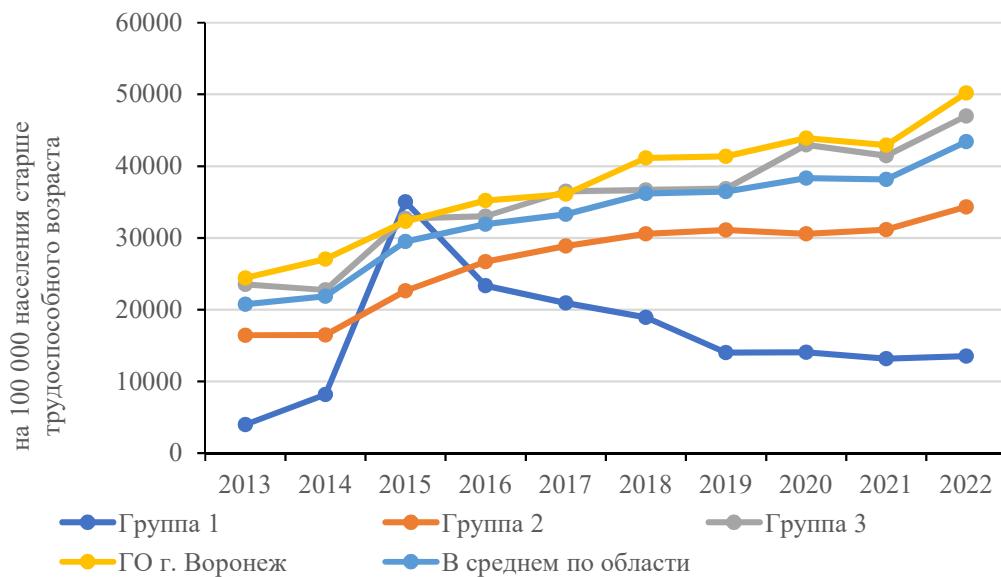


Рисунок 4.3 – Уровень общей заболеваемости болезнями, характеризующимися повышенным кровяным давлением среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

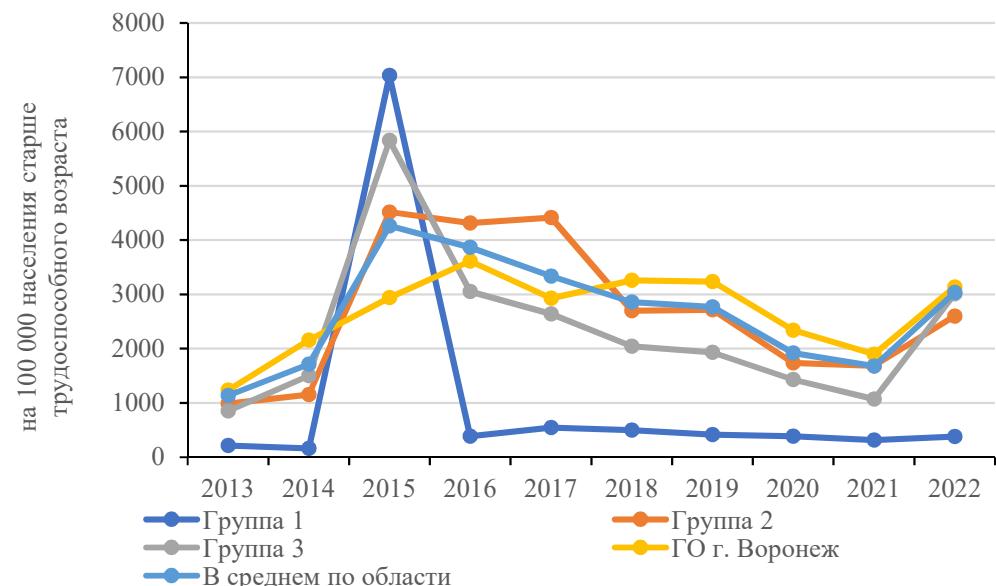


Рисунок 4.4 – Уровень первичной заболеваемости болезнями, характеризующимися повышенным кровяным давлением среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

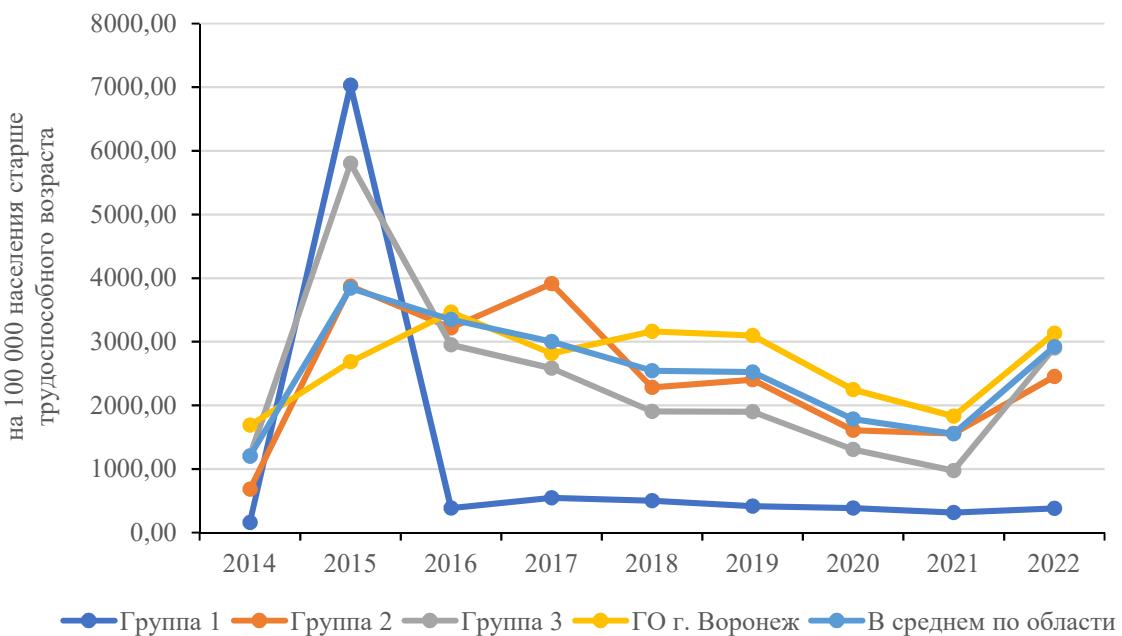


Рисунок 4.5 – Число лиц, взятых под диспансерное наблюдение среди больных с впервые в жизни установленным диагнозом нозологии болезней, характеризующихся повышенным кровяным давлением среди населения Воронежской области по данным за 2014-2022 гг. (на 100 000 населения старше трудоспособного возраста)

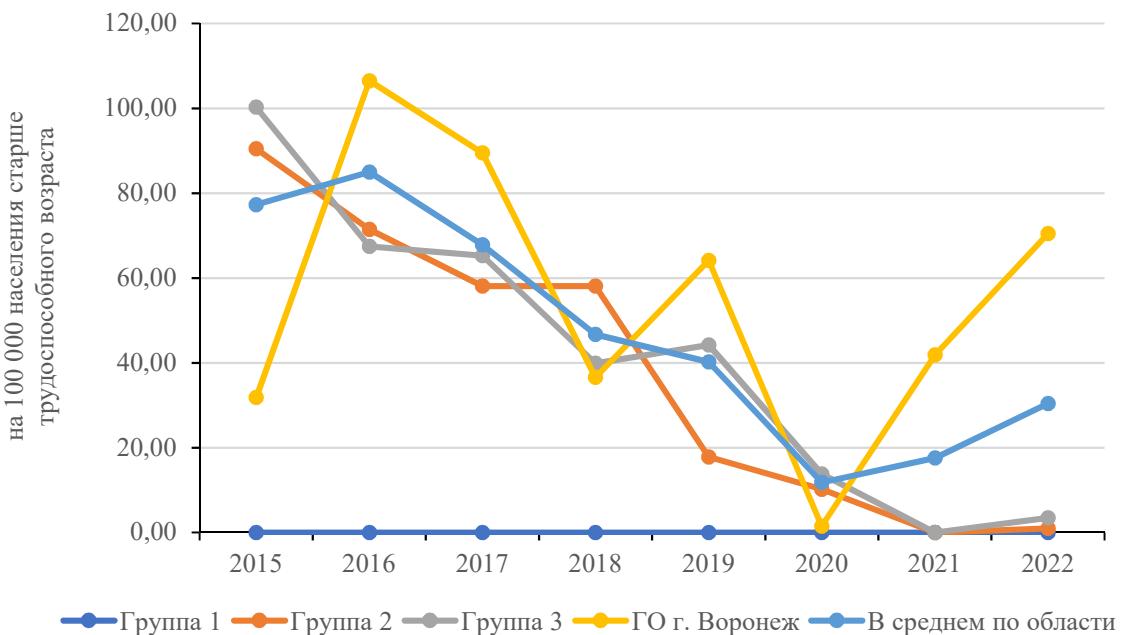


Рисунок 4.6 – Число лиц с впервые установленным диагнозом нозологии болезней, характеризующихся повышенным кровяным давлением, выявленным при профосмотре, среди населения Воронежской области по данным за 2015-2022 гг. (на 100 000 населения старше трудоспособного возраста)

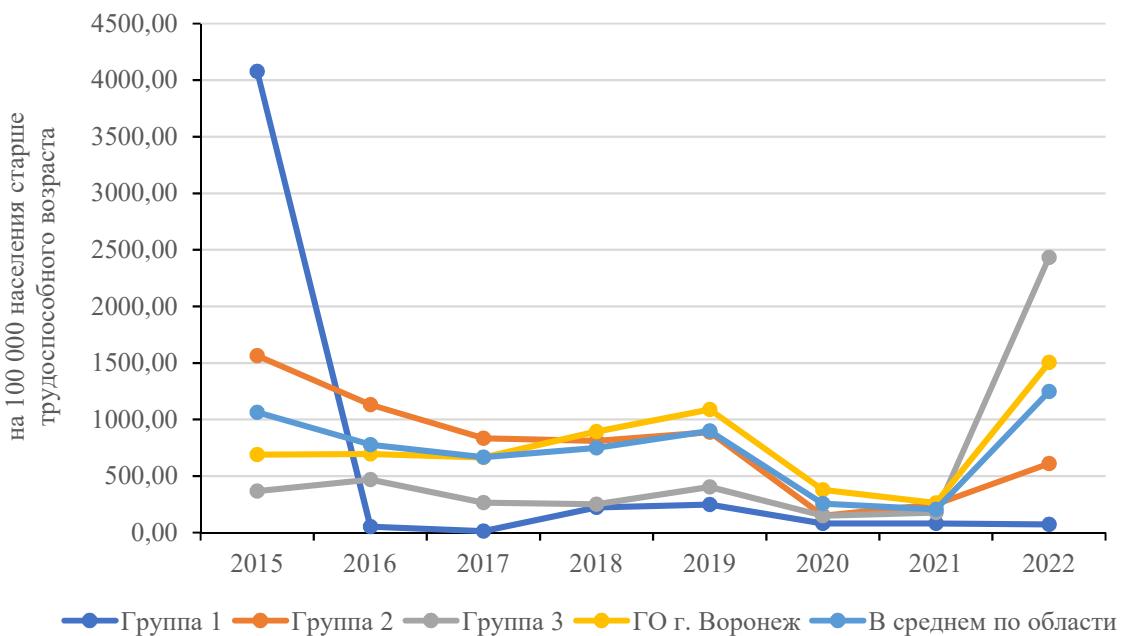


Рисунок 4.7 – Число лиц с впервые установленным диагнозом нозологии болезней, характеризующихся повышенным кровяным давлением, выявленным при диспансеризации, среди населения Воронежской области по данным за 2015-2022 гг. (на 100 000 населения старше трудоспособного возраста)

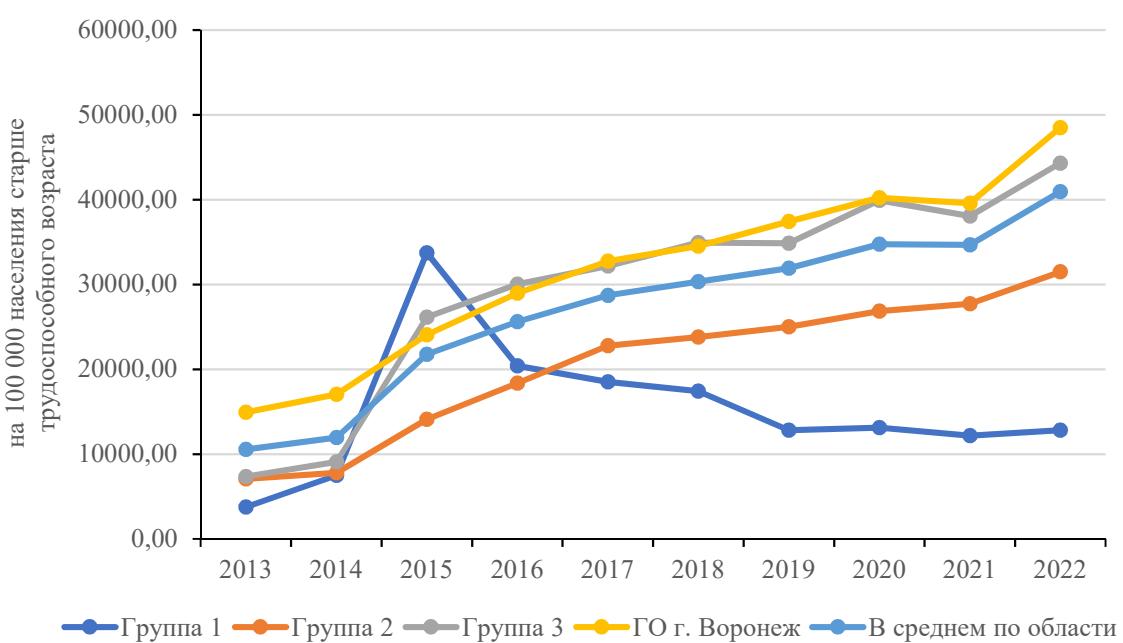


Рисунок 4.8 – Число лиц, страдающих болезнями, характеризующимися повышенным кровяным давлением, состоящих под диспансерным наблюдением на конец года среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

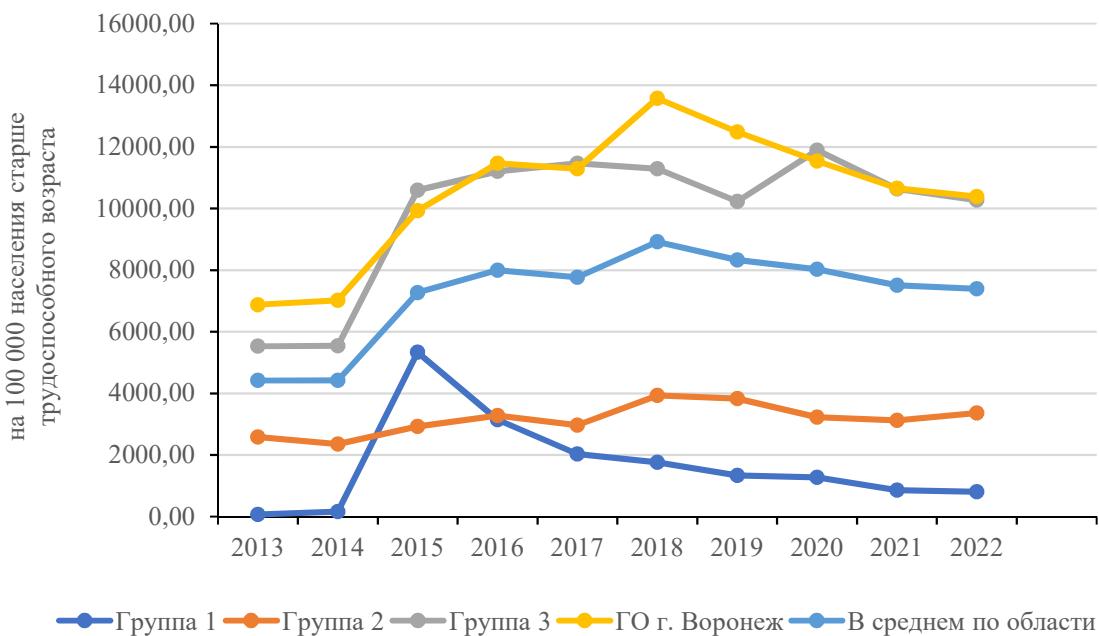


Рисунок 4.9 – Уровень общей заболеваемости эссенциальной гипертензией среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

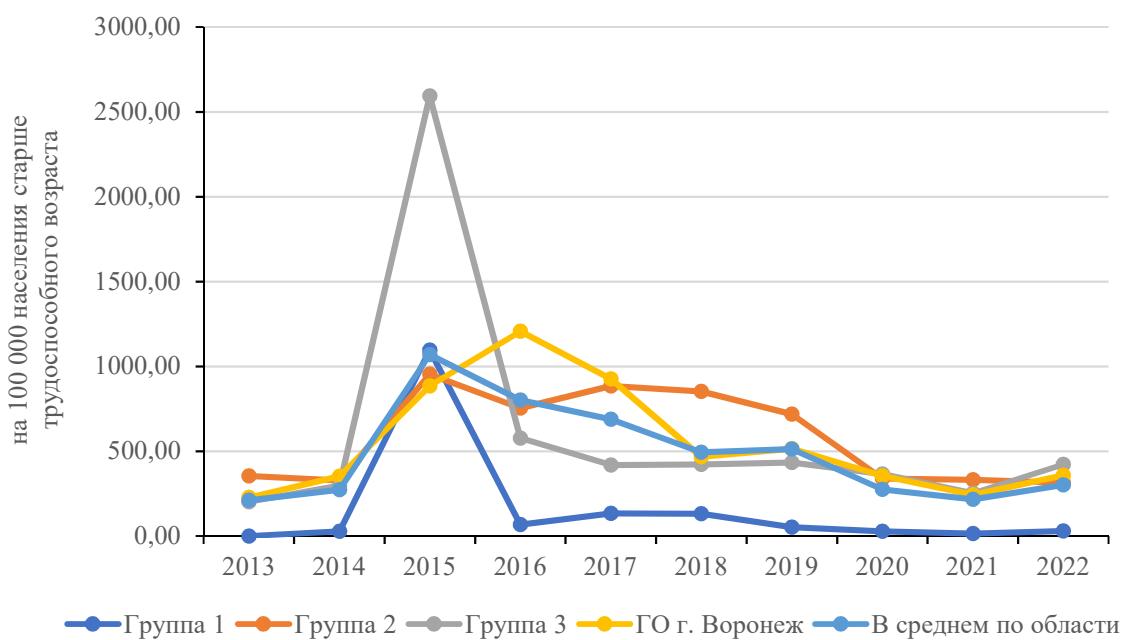


Рисунок 4.10 – Уровень первичной заболеваемости эссенциальной гипертензией среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

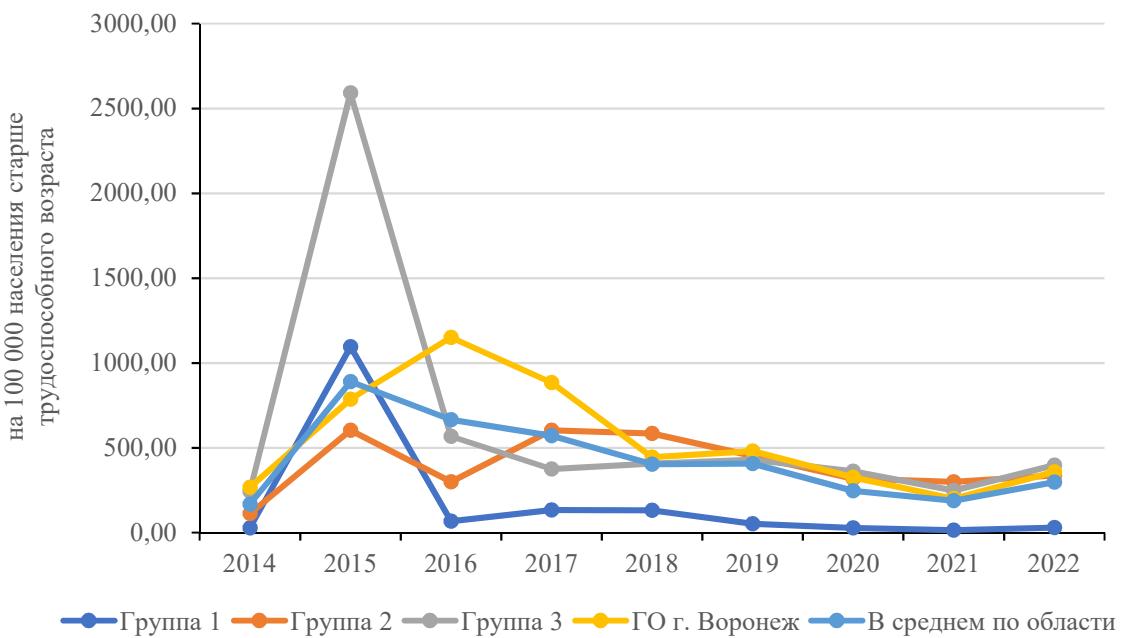


Рисунок 4.11 – Число лиц, взятых под диспансерное наблюдение среди больных с впервые в жизни установленным диагнозом «эссенциальная гипертензия» среди населения Воронежской области по данным за 2014-2022 гг. (на 100 000 населения старше трудоспособного возраста)

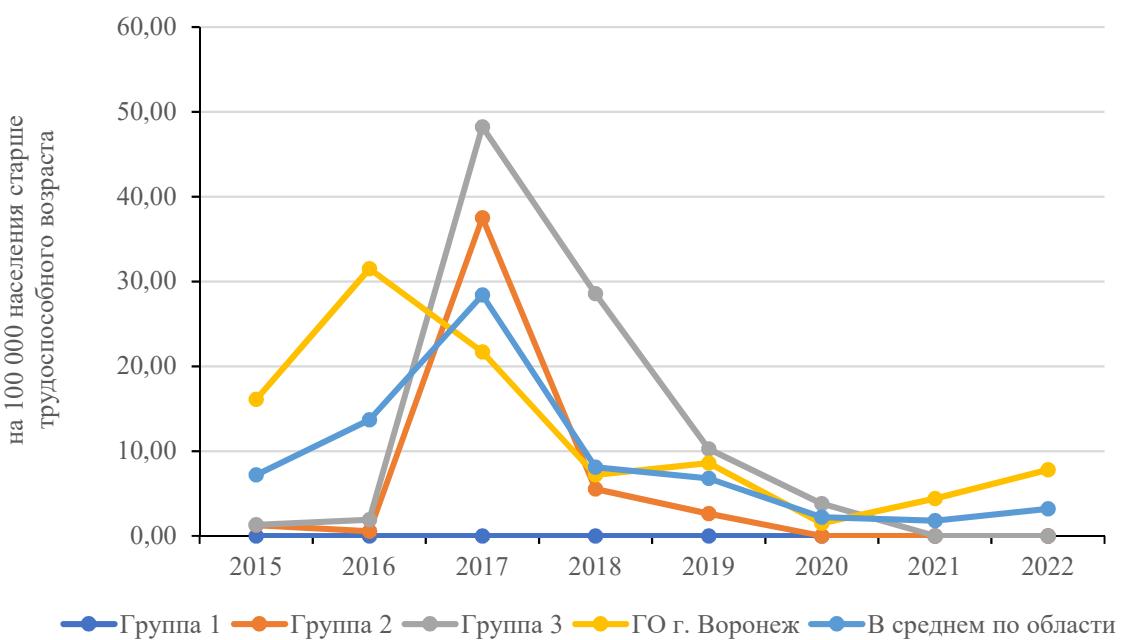


Рисунок 4.12 – Число лиц с впервые установленным диагнозом «эссенциальная гипертензия», выявленным при профосмотре, среди населения Воронежской области по данным за 2015-2022 гг. (на 100 000 населения старше трудоспособного возраста)

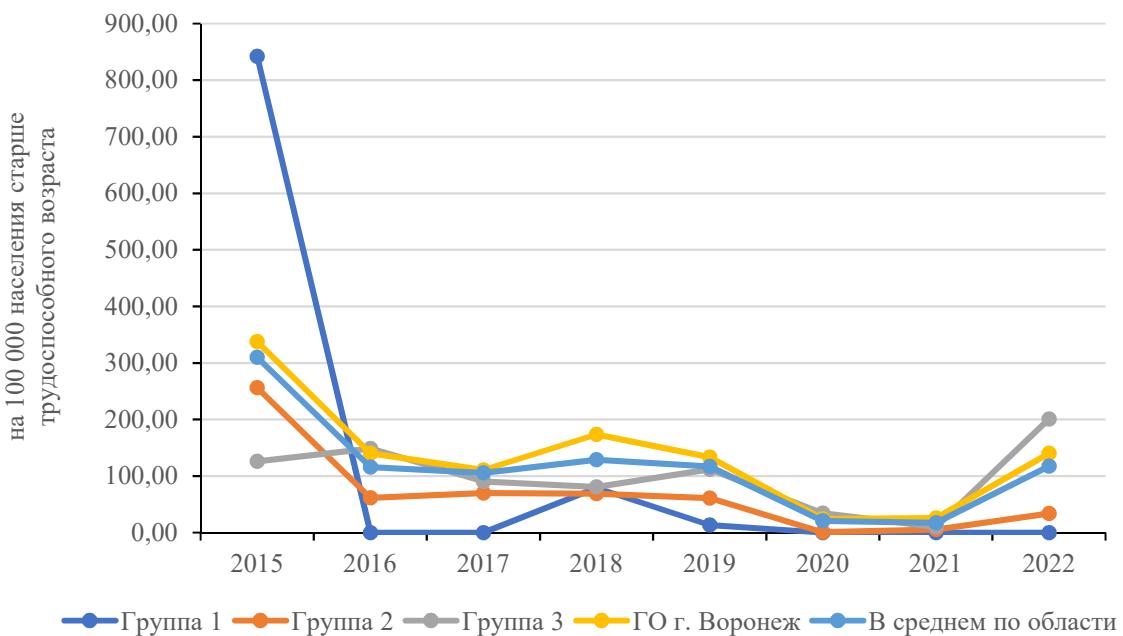


Рисунок 4.13 – Число лиц с впервые установленным диагнозом «эссенциальная гипертензия», выявленным при диспансеризации, среди населения Воронежской области по данным за 2015-2022 гг. (на 100 000 населения старше трудоспособного возраста)

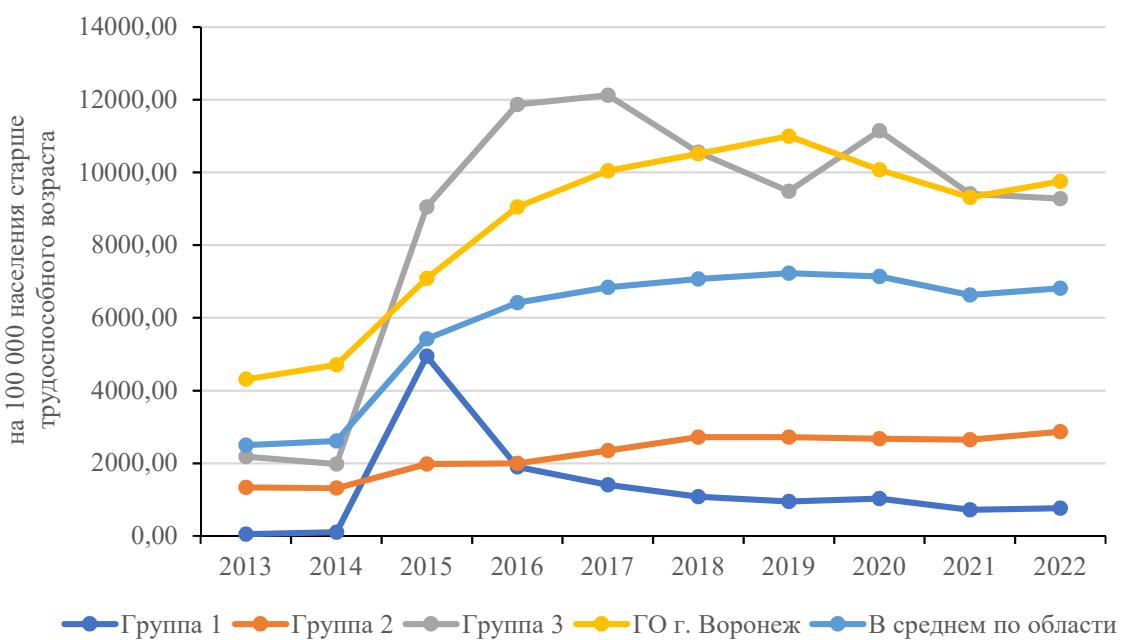


Рисунок 4.14 – Число лиц, страдающих эссенциальной гипертензией, состоящих под диспансерным наблюдением на конец года среди населения Воронежской области по данным за 2013-2022 гг. (на 100 000 населения старше трудоспособного возраста)

Сопоставление результатов визуализации по нозологии $n = 1$ показывает, что у взрослого населения группы 2, группы 3 и областного центра сохраняется тенденция ежегодного прироста уровня общей заболеваемости, а группы 1 заметно снижение, которое, тем не менее, замедляется. Следует также обратить внимание на то, что общая заболеваемость в областном центре выше, чем в группе 3 районов области. Также заметно, что, если темпы прироста снижались (рис. 4.3), то в 2022 году произошло повышение интенсивности эпидемиологического процесса. Характер временного ряда уровня первичной заболеваемости (рис. 4.4) имеет отличия: нерегулярные изменения до 2019 года, уменьшение уровня первичной заболеваемости в период 2019-2021, а после 2021 года совпадает с тенденцией изменения общей заболеваемости по всем территориальным образованиям.

Визуальный анализ позволяет оценить степень активности причин и условий, приводящих к повышению артериального давления у лиц старшего и пожилого возраста. Судя по графикам временных рядов общей заболеваемости (рис. 4.3), их действие включает периоды роста (2013-2017 гг.), стабильности (2018-2021 гг.) и последующего роста в 2022 году. По первичной заболеваемости (рис. 4.4) – периоды нерегулярных изменений (2013-2017 гг.), стабильного снижения (2018-2021 гг.) и последующего роста в 2022 году по всем территориальным образованиям, за исключением группы 1 районов области. Эти оценки позволяют считать, что мероприятия, проводимые в Воронежской области по снижению влияния на заболеваемость нозологией $n = 1$ взрослого населения в силу бытовых, социально-экономических, экологических и других причин, и, соответственно, ресурсное обеспечение, выделяемое органами управления для этих целей, носят недостаточно планомерный и устойчивый характер.

Далее проведена оценка адекватности процесса ресурсного обеспечения – объема диспансеризации взрослого населения Воронежской области тенденциям многолетней динамики заболеваемости на основе графического представления временного ряда $R_1(t_1)$ – число лиц, состоящих под диспансерным наблюдением на конец года (рис. 4.8).

Сопоставление графика временных рядов на рисунках 4.8 и 4.5 характеризуется совпадающими изменениями и тенденциями в многолетней динамике общей заболеваемости и числа лиц, состоящих под диспансерным наблюдением на конец года. Поэтому можно считать, что диспансерное наблюдение лиц старшего и пожилого возраста адекватно отражает интенсивность эпидемиологического процесса по нозологии $n = 1$.

Приведенные оценки результатов визуального моделирования позволяют сделать вывод о дальнейшем продолжении предиктивного анализа на основе прогностического моделирования временных рядов для интеграции этих результатов в процесс принятия управленческих решений.

На основе ансамбля прогностических моделей (таблица 2.7) с наилучшим качеством по точности осуществлено прогностическое моделирование временных рядов с горизонтом планирования $T = 4$. Результаты прогнозов показателя общей заболеваемости $y_1(t)$ и объемов диспансеризации $R_1(t)$ нозологии $n = 1$ представлены в таблице 4.1.

Таблица 4.1 – Результаты прогнозирования временных рядов с горизонтом планирования 4 года

Показатель	Группа выборки	Модель	Прогноз на будущие периоды			
			2023 г.	2024 г.	2025 г.	2026 г.
Общая заболеваемость	Группа 1	MLP	13633,53	12703,36	12852,62	12380,64
	Группа 2	ARIMA	35513,98	36592,14	37575,59	38472,67
	Группа 3	RF	48742,64	50975,8	53491,13	54242,9
	ГО г. Воронеж	HW	51616,13	54189,91	56763,7	59337,49
	Область	HW	45610,99	47907,47	50203,95	52500,44
Объем диспансеризации	Группа 1	DT	13429,3	14052,7	14676,1	15299,5
	Группа 2	ARIMA	34143,93	36792,1	39440,27	42088,44
	Группа 3	RF	45524,77	49251,81	50149,21	53876,25
	ГО г. Воронеж	ARIMA	52140,94	55785,57	59429,9	63073,91
	Область	HW	44393,54	44393,54	44393,54	44393,54

Использование результатов предиктивного анализа для получения управленческих решений на основе оптимизации позволит определить распределение $\Delta R_{nd}(t)$, $n = \overline{1,2}; d = \overline{1,5}$, по временным периодам $t = \overline{10 + 1, 10 + 4}$ с горизонтом прогнозирования $T = 4$.

Полученные результаты прогнозирования временных рядов позволяют сформировать модель многоальтернативной оптимизации с учетом выбора экспертом значений термов лингвистической переменной для каждого временного периода горизонта планирования для принятия управленческого решения при распределении ресурсного обеспечения, выраженного в объемах диспансеризации, между группами населения региона. Из полученного объема дополнительного ресурса в рамках каждой группы населения аналогичным образом осуществляется его распределение по территориальным образованиям путем формирования оптимизационной модели и принятия управленческого решения. Затем осуществляется коррекция распределения дополнительного ресурса по временным периодам путем формирования единой целевой функции максимизации для всего горизонта планирования на основе прогностической модели интегрального эффекта, используемой в окончательном формировании многомерной оптимизационной модели, для поиска окончательных управленческих решений.

4.3. Оценка эффективности применения разработанных средств при управлении процессом диспансеризации на основе вычислительного эксперимента

Проведены исследование и выявление наилучшей по качеству модели регрессии, построенной на разных вариантах наборов данных на основе массивов (2.5)-(2.7) из п. 2.2, для дальнейшего прогнозирования значений показателей нозологии на основании ретроспективной информации предшествующих периодов заболеваемости по группе заболеваний – артериальная гипертензия лиц старших возрастных групп и их диспансеризации. Решены следующие задачи.

1. Формирование трех вариантов наборов данных для предиктивного анализа и моделирования.

2. Предварительная подготовка сформированных наборов данных и исследование корреляции между показателями нозологий.

3. Проведение обучения с использованием подготовленных вариантов наборов данных, оценка качества моделей регрессии и сравнение полученных результатов.

Для каждой из представленных задач сформированы варианты обучающих наборов данных. Для этого использованы данные статистики заболеваемости Воронежской области по группе болезней – артериальная гипертензия за 10-летний период с 2013 по 2022 годы.

Для регрессионного моделирования было сформировано три набора данных:

- данные одного показателя одной нозологии в рамках каждого района отдельно (задача 1);
- данные одного показателя одной нозологии с учетом разных районов (задача 2);
- данные показателей всех нозологий предыдущего периода в рамках каждого района отдельно (задача 3).

Данные одного показателя (общая заболеваемость) одной нозологии (болезни, характеризующиеся повышенным кровяным давлением) в рамках каждого района отдельно содержат следующие атрибуты:

- District – строковое обозначение районов Воронежской области (тип данных object);
- Year – год прогноза временного ряда (тип данных int64);
- Col1-Col3 – значения показателя за предыдущие 3 года (типы данных float64);
- Forecast – значение показателя в год прогнозирования (тип данных int64).

В таблице 4.2 представлена структура сформированного набора данных для решения задачи (1).

Таблица 4.2 – Структура сформированного набора данных для варианта: один показатель – одна нозология в рамках каждого района отдельно

Район	Год	Значения показателя за предыдущие 3 года			Прогноз
District	Year	Col1	Col2	Col3	Forecast
Аннинский	2016	11792,2	15480,0	21425,4	27027,6
Аннинский	2017	15480,0	21425,4	27027,6	27026,8
Аннинский	2018	21425,4	27027,6	27026,8	35043,2
Аннинский	2019	27027,6	27026,8	35043,2	31947,8
Аннинский	2020	27026,8	35043,2	31947,8	32903,2
Аннинский	2021	35043,2	31947,8	32903,2	45673,4
Аннинский	2022	31947,8	32903,2	45673,4	45137,2
Бобровский	2016	29318,4	23983,0	24950,2	30982,8
Бобровский	2017	23983,0	24950,2	30982,8	37186,1
Бобровский	2018	24950,2	30982,8	37186,1	38368,3
Бобровский	2019	30982,8	37186,1	38368,3	39126,9
Бобровский	2020	37186,1	38368,3	39126,9	41032,2
Бобровский	2021	38368,3	39126,9	41032,2	39498,8
Бобровский	2022	39126,9	41032,2	39498,8	45141,5
...

Обучающий набор данных содержит 231 строку и 6 столбцов.

Данные одного показателя одной нозологии с учетом разных районов, помимо описанных для первого варианта набора данных, включают столбцы AddCol, пронумерованных от 1 до 96, которые содержат значения показателя за предыдущие 3 года по разным районам (типы данных float64). В таблице 4.3 представлена структура сформированного набора данных, который содержит 231 строку и 102 столбца, и предлагается для использования при решении задачи (2).

Таблица 4.3 – Структура сформированного набора данных для варианта: один показатель – одна нозология с учетом разных районов

Район	Год	Значения показателя за предыдущие 3 года			Значения показателя за предыдущие 3 года по разным районам				Прогноз
		Col1	Col2	Col3	AddCol 1	AddCol 2	AddCol 3	...	
District	Year	Col1	Col2	Col3	AddCol 1	AddCol 2	AddCol 3	...	Forecast
Аннинский	2016	11792,2	15480,0	21425,4	29318,4	23983,0	24950,2	...	27027,6
Аннинский	2017	15480,0	21425,4	27027,6	23983,0	24950,2	30982,8	...	27026,8
Аннинский	2018	21425,4	27027,6	27026,8	24950,2	30982,8	37186,1	...	35043,2
Аннинский	2019	27027,6	27026,8	35043,2	30982,8	37186,1	38368,3	...	31947,8
Аннинский	2020	27026,8	35043,2	31947,8	37186,1	38368,3	39126,9	...	32903,2
Аннинский	2021	35043,2	31947,8	32903,2	38368,3	39126,9	41032,2	...	45673,4
Аннинский	2022	31947,8	32903,2	45673,4	39126,9	41032,2	39498,8	...	45137,2
Бобровский	2016	29318,4	23983,0	24950,2	11792,2	15480,0	21425,4	...	30982,8
Бобровский	2017	23983,0	24950,2	30982,8	15480,0	21425,4	27027,6	...	37186,1
Бобровский	2018	24950,2	30982,8	37186,1	21425,4	27027,6	27026,8	...	38368,3
Бобровский	2019	30982,8	37186,1	38368,3	27027,6	27026,8	35043,2	...	39126,9
Бобровский	2020	37186,1	38368,3	39126,9	27026,8	35043,2	31947,8	...	41032,2
Бобровский	2021	38368,3	39126,9	41032,2	35043,2	31947,8	32903,2	...	39498,8
Бобровский	2022	39126,9	41032,2	39498,8	31947,8	32903,2	45673,4	...	45141,5
...

Данные показателей всех нозологий предыдущего периода в рамках каждого района отдельно, помимо описанных для первого варианта набора данных, включают столбцы AddCol с указанием номера показателя (Р, от 2 до 6) и номера нозологии (N, от 1 до 5), пронумерованных от 1 до 87, которые содержат значения показателей по всем нозологиям за предыдущие 3 года по разным районам (типы данных float64). В таблице 4.4 представлена структура сформированного набора данных для решения задачи (3), который содержит 231 строку и 93 столбца.

Таблица 4.4 – Структура сформированного набора данных для варианта: все показатели – все нозологии в рамках каждого района отдельно

Район District	Год Year	Значения показателя за предыдущие 3 года			Значения показателей по всем нозологиям				Прогноз Forecast
		Col1	Col2	Col3	AddCol 1_P2_N 1	AddCol 2_P2_N 1	AddCol 3_P2_N 1	...	
Аннинский	2016	11792,2	15480,0	21425,4	698,5	564,3	4958,6	...	27027,6
Аннинский	2017	15480,0	21425,4	27027,6	564,3	4958,6	6011,0	...	27026,8
Аннинский	2018	21425,4	27027,6	27026,8	4958,6	6011,0	2572,6	...	35043,2
Аннинский	2019	27027,6	27026,8	35043,2	6011,0	2572,6	3336,7	...	31947,8
Аннинский	2020	27026,8	35043,2	31947,8	2572,6	3336,7	6278,7	...	32903,2
Аннинский	2021	35043,2	31947,8	32903,2	3336,7	6278,7	1669,8	...	45673,4
Аннинский	2022	31947,8	32903,2	45673,4	6278,7	1669,8	1730,6	...	45137,2
Бобровский	2016	29318,4	23983,0	24950,2	705,3	1126,6	1568,1	...	30982,8
Бобровский	2017	23983,0	24950,2	30982,8	1126,6	1568,1	16107,7	...	37186,1
Бобровский	2018	24950,2	30982,8	37186,1	1568,1	16107,7	9622,4	...	38368,3
Бобровский	2019	30982,8	37186,1	38368,3	16107,7	9622,4	3327,0	...	39126,9
Бобровский	2020	37186,1	38368,3	39126,9	9622,4	3327,0	3538,0	...	41032,2
Бобровский	2021	38368,3	39126,9	41032,2	3327,0	3538,0	2307,0	...	39498,8
Бобровский	2022	39126,9	41032,2	39498,8	3538,0	2307,0	1615,6	...	45141,5
...

Перед началом проведения обучения регрессионных моделей была проведена предварительная подготовка данных, которая необходима для совместимости с требованиями алгоритмов регрессии, а также для снижения вероятности влияния на результаты корреляционного анализа.

Для третьего сформированного варианта набора данных характерно наличие пропусков, связанных с отсутствием информации за 2013 и 2014 годы по показателям 3-5 в нозологиях. Было принято решение заполнить данные пропуски медианными значениями в разрезе атрибута «District».

Следующим этапом подготовки данных является обработка категориальных признаков. В наборах данных присутствует категориальный атрибут «District». Он относится к номинальному виду, что означает невозможность проранжирования его в целях упорядочивания, поэтому данный атрибут был преобразован в набор числовых признаков с использованием метода Binary Encoding («Двоичное кодирование») для каждого исследуемого варианта.

При исследовании корреляции в многомерных наборах данных используется корреляционная матрица. Множественная корреляция измеряет степень зависимости одной переменной от нескольких других и является расширением понятия простой корреляции [92].

Для обучающего набора с данными одного показателя одной нозологии в рамках каждого района отдельно построена корреляционная матрица, представленная на рисунке 4.15. По ней можно заметить, что для целевого атрибута «Forecast» сильную положительную корреляцию имеют данные за прошлый год (Col3), при этом за следующие два года прошлого периода положительная корреляция ниже, но остается заметно высокой.



Рисунок 4.15 – Корреляционная матрица для обучающего набора с данными одного показателя одной нозологии в рамках каждого района отдельно

Для обучающего набора с данными одного показателя одной нозологии с учетом разных районов строились корреляционные матрицы для каждого из трех атрибутов, содержащие информацию за предыдущие годы в других районах региона в целях удобного визуального анализа ввиду большого количества столбцов. Было обнаружено, что значения показателя общей заболеваемости по

другим районам имеют слабую положительную корреляцию (ближе к нулю) относительного целевого атрибута «Forecast».

Для обучающего набора с данными показателей всех нозологий предыдущего периода в рамках каждого района отдельно также строились корреляционные матрицы для каждого трех атрибутов, содержащие информацию за предыдущие годы по показателям всех нозологий. Была выявлена средняя положительная корреляция по следующим показателям и нозологиям относительного целевого атрибута «Forecast»:

- по показателю 6 нозологии 1 (рисунок 4.16);
- по показателю 1 нозологии 3 (рисунок 4.17);
- по показателю 6 нозологии 3 (рисунок 4.18).

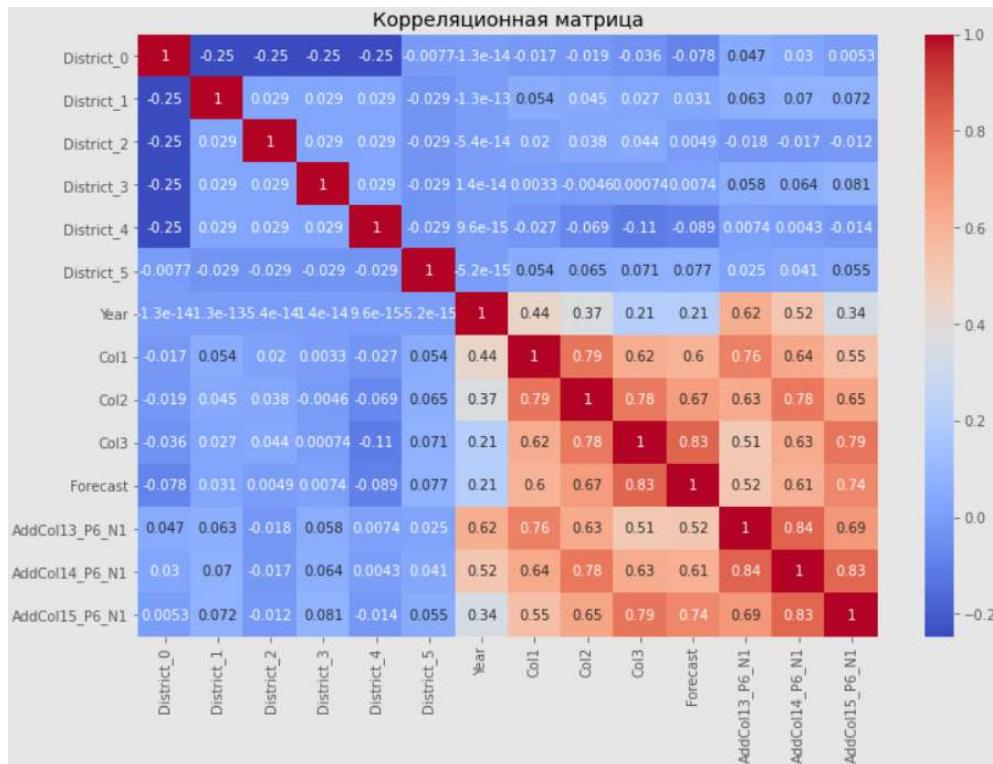


Рисунок 4.16 – Корреляционная матрица для обучающего набора с данными показателей всех нозологий предыдущего периода в рамках каждого района отдельно (показатель 6, нозология 1)

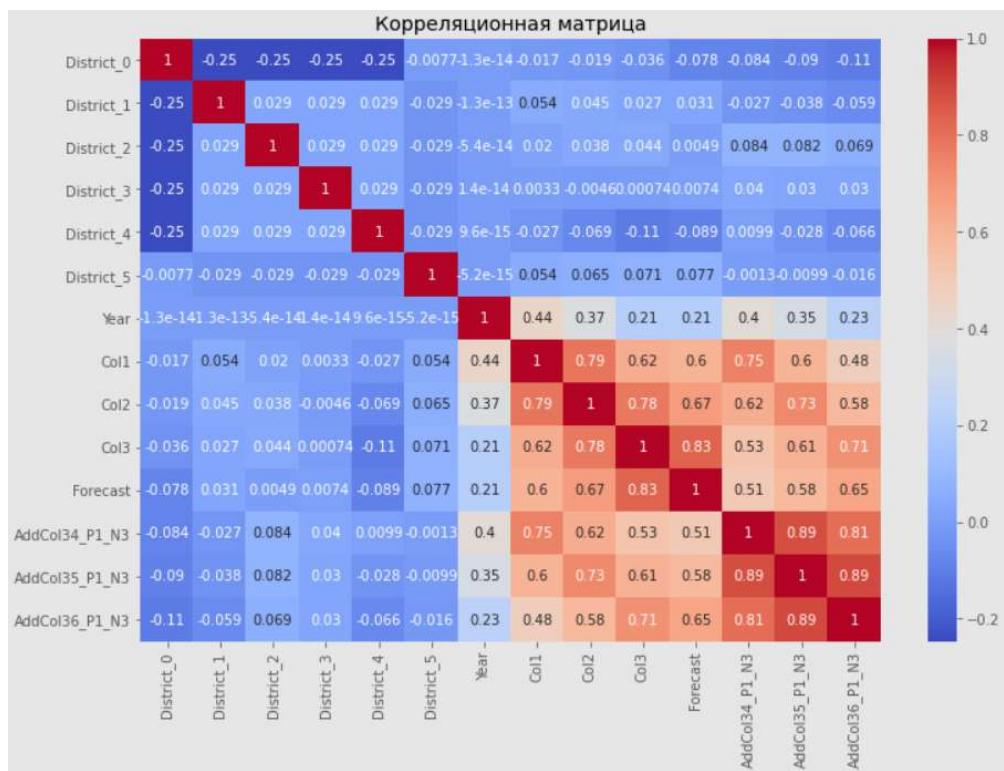


Рисунок 4.17 – Корреляционная матрица для обучающего набора с данными показателей всех нозологий предыдущего периода в рамках каждого района отдельно (показатель 1, нозология 3)

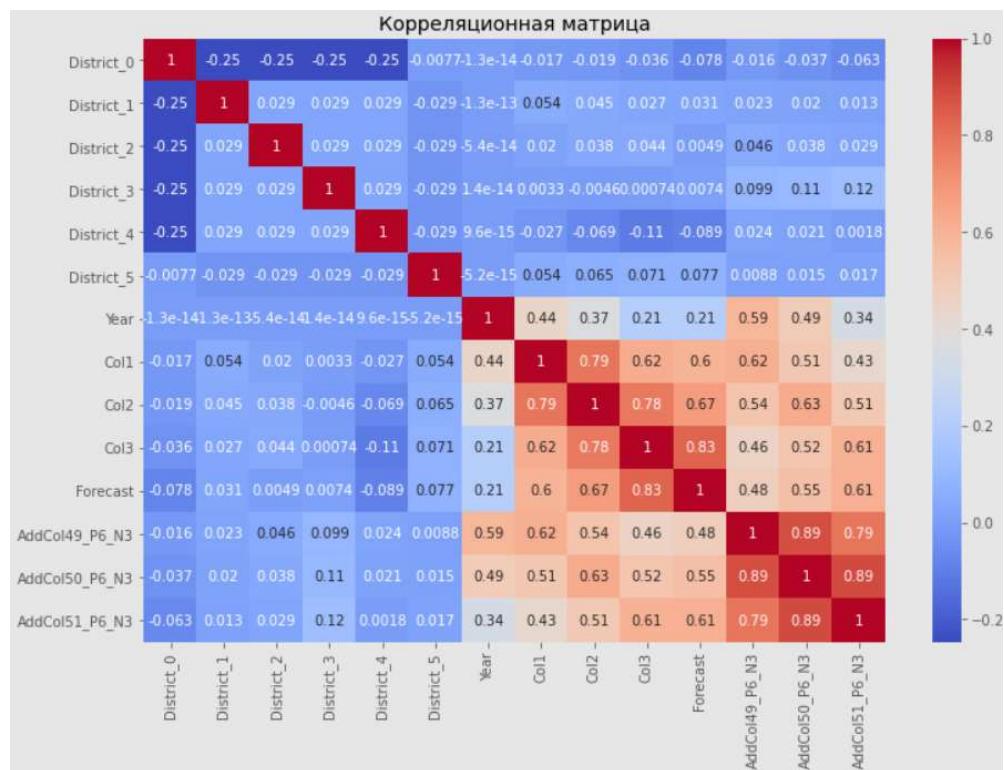


Рисунок 4.18 – Корреляционная матрица для обучающего набора с данными показателей всех нозологий предыдущего периода в рамках каждого района отдельно (показатель 6, нозология 3)

По другим показателям других нозологий выявлены слабые положительные корреляции.

Регрессионное моделирование и сравнение результатов осуществляется с использованием следующих алгоритмов машинного обучения:

- моделирование с использованием деревьев решений (Decision tree regressor, DTR) и случайного леса (Random forest regressor, RFR).
- нейросетевое моделирование с использованием многослойного перцептрона (Multi-layer Perceptron regressor, MLPR).

Ошибка прогнозирования каждого из используемых методов оценивалась с использованием показателей среднеквадратичной ошибки RMSE, средней абсолютной ошибки MAE и коэффициента детерминации R2.

При проведении регрессионного моделирования наборы данных были разделены на обучающую и тестовую выборки в соотношении 87/13. В таблице 4.5 представлены сравнительные результаты работы моделей регрессии.

Таблица 4.5 – Сравнение результатов работы моделей регрессии

Вариант набора данных	Регрессор	Метрики качества		
		RMSE	MAE	R2
Данные одного показателя одной нозологии в рамках каждого района отдельно	DTR	8986.88	6430.02	0.56
	RFR	9804.45	6588.85	0.48
	MLPR	7122.59	5091.21	0.72
Данные одного показателя одной нозологии с учетом разных районов	DTR	8983.88	6557.47	0.56
	RFR	9099.32	5985.18	0.55
	MLPR	8827.5	5993.58	0.58
Данные показателей всех нозологий предыдущего периода в рамках каждого района отдельно	DTR	9712.13	6928.32	0.49
	RFR	9756.67	6487.64	0.48
	MLPR	8268.83	6264.32	0.63

Серым цветом отмечены методы, в которых зафиксировано наилучшее качество прогнозирования в разрезе группы выборки по каждому из показателей для каждой нозологии. Можно заметить, что модель регрессии, построенная на

основе алгоритма нейронной сети многослойный перцепtron, показала наилучший результат во всех трех исследуемых вариантах наборов данных.

Полученные модели на основе алгоритма MLPR можно использовать для прогнозирования временных рядов на будущие периоды для принятия управленческих решений изменения объемов выделяемого ресурсного обеспечения для территориальных образований региональной организационной системы здравоохранения.

В целях повышения эффективности поддержки принятия управленческих решений при анализе ретроспективной информации в региональных отраслевых организационных системах могут быть использованы результаты регрессионного и корреляционного моделирования для различных вариантов наборов данных, которые позволяют анализировать степень влияния разных территориальных образований региона, а также разных показателей эффективности, объединенных в группу по определенному признаку, на социальные показатели, в частности, уровень заболеваемости, в отдельно взятом районе региона.

С целью повышения точности моделей регрессии следует осуществлять предварительную обработку данных и выбирать метод прогнозирования временных рядов, основанный на алгоритмах регрессии, с наилучшим качеством прогноза на заданный временной период.

Выводы четвертой главы

1. Для того, что апробировать разработанные средства предиктивного анализа и оптимизации в практике управления региональной организационной системой, разработано программное обеспечение, объединяющие оригинальные

средства и стандартные программные продукты машинного обучения в рамках структурной модели интеллектуальной поддержки принятия решений.

2. С целью отработки применения средств первого уровня схемы управления диспансеризацией населения старших возрастных групп в региональной системе здравоохранения приемлемо использовать данные многолетней статистики по показателям заболеваемости артериальной гипертензией и объемам диспансерного учета, детализованные по группам населения и территориальным образованиям Воронежской области. На основании этих массивов временных рядов осуществляется дообучение прогностических моделей с использованием моделей, предобученных по данным тестовой выборки.

3. Оценку эффективности разработанных средств интеллектуализации первого и второго уровня целесообразно осуществлять путем сравнения результатов вычислительного эксперимента при принятии управленческих решений по распределению объемов диспансеризации и результатов экспертных решений, отраженных в прогностических моделях, обученных с использованием ретроспективной информации. Сравнение приемлемо проводить по величине интегрального эффекта и времени, затрачиваемого на принятие решений.

ЗАКЛЮЧЕНИЕ

Учитывая ориентацию на использование данных многолетнего статистического учета, региональные отраслевые организационные системы приобретают ряд особенностей при управлении распределением ресурсного обеспечения. Поэтому требуются исследования, развивающие методы и механизмы принятия управленческих решений, которые являются инвариантными в теории управления.

Прежде всего особенности этого класса организационных систем проявляются, во-первых, в структурном моделировании взаимодействия управляющего центра и компонентов системы, а во-вторых, при поддержке решений управляющего центра с использованием двухуровневой схемы, обеспечивающей интеграцию результатов предиктивного анализа в процесс выработки решений на основе оптимизации. Эти модели определяют пути повышения эффективности управления в региональных отраслевых организационных системах с использованием ретроспективной информации.

Структурные модели служат базовой основой для последующей разработки проблемно-ориентированных средств предиктивного анализа с формированием результатов в виде ансамбля предобученных прогностических моделей, на основе которых получают прогностические модели, необходимые в процессе формирования оптимизационных моделей.

Необходимость распределения дополнительного ресурсного обеспечения, с одной стороны, между группами населения и территориальными образованиями региона для фиксированных временных периодов в рамках горизонта планирования, а с другой – одновременно для всех временных периодов, приводят к разработкам комплекса процедур принятия управленческих решений, отличающихся разными формами оптимизационного моделирования.

В результате исследований:

1. Проанализированы пути повышения эффективности управления в региональных отраслевых организационных системах с использованием ретроспективной информации за счет применения предиктивного анализа статистических данных и интеграции его результатов в процесс принятия решений на основе оптимизации.

2. Сформирована структурная модель управления ресурсным обеспечением в региональной отраслевой организационной системе, включающая компоненты двухуровневой схемы процесса принятия решений.

3. Предложена алгоритмическая схема предиктивного анализа ретроспективной информации в региональной отраслевой организационной системе, объединяющая экспертное моделирование, моделирование на основе машинного обучения и выбор наилучшего по точности метода прогнозирования путем сравнительного анализа результатов обработки тестовой выборки.

4. Разработана процедура интеграции результатов предиктивного анализа ретроспективной информации в процесс принятия управленческих решений по распределению ресурсного обеспечения между группами населения региона.

5. Проведена детализация управленческого решения для групп населения региона на основе процедуры интеграции результатов предиктивного анализа в процесс многоальтернативной оптимизации при распределении ресурсного обеспечения по территориальным образованиям.

6. Обоснована необходимость коррекции детализированного распределения ресурсного обеспечения путем интеграции результатов предиктивного анализа в процесс принятия управленческого решения на базе полностью формализованного описания экстремального и граничных требований оптимизации объектов ресурса по временным периодам в рамках заданного горизонта планирования.

7. Сформирована структура программного обеспечения системы управления ресурсным обеспечением на основе ретроспективной информации, фиксируемой в региональной отраслевой организационной системе.

8. Проведена оценка применения разработанных средств в практике управления ресурсным обеспечением диспансеризации населения в региональной системе здравоохранения Воронежской области. Получены прогностические оценки при интеграции результатов предиктивного анализа в принятии решений на основе оптимизации, характеризующие рост интегрального эффекта на 12%, сокращение времени на выработку управленческого решения на 58%.

Рекомендации и перспективы дальнейшей разработки темы

1. Результаты диссертационного исследования рекомендуются к применению в задачах управления распределением ресурсного обеспечения в региональных отраслевых организационных системах, отличительной особенностью которых является принятие решения по трем составляющим распределения ресурса: группам населения региона, территориальным образованиям и временным периодам в рамках горизонта планирования на основе ретроспективной информации.

2. Дальнейшая разработка темы будет направлена на исследование возможностей применения элементов нечеткой логики, а также апробации по другим отраслям региональных организационных систем.

СПИСОК ЛИТЕРАТУРЫ

- 1 Кремер Н.Ш. Теория вероятностей и математическая статистика. В 2 частях. Ч. 2. Математическая статистика: учебник и практикум для академического бакалавриата / Н.Ш. Кремер. – 4-е изд., перераб. и доп. – Москва: Издательство Юрайт, 2016.
- 2 Назаров М.Г. Курс социально-экономической статистики: Учебник для ВУЗов / Под ред. Проф. М.Г. Назарова. – М.: Финстатинформ, ЮНИТИ-ДАНА, 2014. – 771с.
- 3 Ефимова М.Р. Социально-экономическая статистика: учебник для академического бакалавриата / М.Р. Ефимова, А.С. Аброскин, С.Г. Бычкова, М.А. Михайлов; под редакцией М. Р. Ефимовой. – 2-е изд., перераб. и доп. – Москва: Издательство Юрайт, 2017. – 591 с.
- 4 Едронова В.Н. Общая теория статистики: учебник / В.Н. Едронова, М.В. Малафеева. – 2-е изд., перераб. и доп. – Москва: Магистр: ИНФРА-М, 2022. – 608 с.
- 5 Горленко О.А. Статистические методы в управлении качеством: учебник и практикум для вузов / О.А. Горленко, Н.М. Борбаць; под редакцией О.А. Горленко. – 2-е изд., испр. и доп. – Москва: Издательство Юрайт, 2023. – 306 с.
- 6 Жгун Т.В. Применение метода конечных разностей для оценки качества данных / Т.В. Жгун, А.А. Жгун // Системный анализ в проектировании и управлении. В 3ч. Ч.2: сборник научных трудов XXV Международной научной и учебно-практической конференции, 13-14 октября 2021 г. – СПб.: ПОЛИТЕХ-ПРЕСС, 2021. – С. 64-71.
- 7 Коротков М.Г. О некоторых подходах преобразования тестовых баллов в отметки / М.Г. Коротков, А.А. Петров // Вестник ЮГУ. 2023. №2 (69). С. 9-16.

8 Терованесов М.Р. Эконометрические модели как инструмент повышения эффективности управления организациями железнодорожного транспорта / М.Р. Терованесов, В.С. Козлов // Сборник научных трудов ДОНИЖТ. 2022. №64. С. 12-20.

9 Лабинский А.Ю. Поддержка принятия управленческих решений при рисках чрезвычайных ситуаций на основе применения методов анализа многомерных статистических данных / А.Ю. Лабинский, И.В. Бородушко, А.А. Таранцев // Инженерный вестник Дона. – 2023. – №10 (106).

10 Богданова Е.М. Информационная система прогнозирования чрезвычайных ситуаций при использовании адаптивных моделей / Е.М. Богданова, А.В. Максимов, А.В. Матвеев // Научно-аналитический журнал "Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России". 2019. № 2. С. 65-70

11 Маркосян З.С. Качественный и достоверный статистический учет в медицинской организации / З.С. Маркосян, В.В. Кожевников, Е.А. Черных, Н.А. Щетинина // International journal of professional science. 2022. №5. С. 115-130.

12 Какорина Е.П. Некоторые проблемы медицинской статистики в Российской Федерации / Е.П. Какорина, Е.В. Огрызко // Менеджер здравоохранения. 2012; 6. С. 40-46.

13 Тараксина А.В., Зурнаджьянц Ю.А. Внутренняя отчетность в системе анализа эффективности деятельности государственных учреждений здравоохранения. Международный бухгалтерский учет. 2016; 3 (393). С. 18-36.

14 Колесник В.И., Торопова И.С. Совершенствование процессов управления здравоохранением в регионе: статистический аспект. Проблемы современной экономики. 2016; 4 (60). С. 200-203.

15 Лазарева М.Л. Финансовая статистическая отчетность медицинских организаций: существующие недостатки и направления оптимизации / М.Л. Лазарева, И.В. Тюрина // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. 2018. №4. С. 60-65.

- 16 Путина С.А. Финансовый мониторинг и контроль реализации территориальных программ государственных гарантий бесплатной медпомощи: актуальные проблемы и направления развития. Фундаментальные исследования. 2015; 8. С. 605-610.
- 17 Вериковская С.В. Актуальность, проблемы, особенности и тенденции развития аутсорсинга в медицинских организациях / С.В. Вериковская, Е.Н. Гордеева // Маркетинг и логистика. 2017; 2 (10). С. 33-44.
- 18 Сафонов С.М. Совершенствование деятельности кабинета медицинской статистики городской поликлиники: автореф. дисс. канд. мед. наук. – СПб.2016, 27 с.
- 19 Царегородцева М.С. Особенности организации работы медицинского статистика амбулаторно-поликлинического учреждения / М.С. Царегородцева, Г.К. Кострицина // Бюллетень медицинских интернет-конференций 2015; 5(5).
- 20 Голубев Н.А. Исторические аспекты методологии сбора и обработки медико-статистической информации в Российской Федерации / Н.А. Голубев, А.В. Поликарпов, Е.В. Огрызко, И.Б. Шикина, О.О. Захарченко // Социальные аспекты здоровья населения. 2022. №5.
- 21 Волкова О.А. Региональные аспекты оценки эффективности управления ресурсами медицинских организаций / О.А. Волкова, Е.В. Смирнова, С.Н. Черкасов // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. 2020. №4. С. 401-412.
- 22 Клименко Ю.А. Анализ возможностей роста эффективности рейтингового управления для организационных систем / Ю.А. Клименко, Ю.П. Преображенский, Э.М. Львович, В.Н. Кострова // Международный научно-исследовательский журнал. 2022. №8 (122).
- 23 Львович Я.Е. Оптимизация распределения ресурсного обеспечения развития цифровой среды управления в организационных системах / Я.Е. Львович, Н.А. Рындин, Ю.С. Сахаров // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. – 2021. – № 4. – С. 106-114.

24 Львович Я.Е. Алгоритм оптимального распределения ресурсного обеспечения на стадии развития и функционирования цифровой среды управления / Я.Е. Львович, Н.А. Рындин // Системы управления и информационные технологии. – 2022. – № 3(89). – С.54-61.

25 Рындин Н.А. Модели и алгоритмы оптимизации ресурсного обеспечения процесса функционирования цифровой среды управления в организационных системах / Н. А. Рындин // Оптимизация и моделирование в автоматизированных системах: труды Международной молодежной школы. – Воронеж. – 2022. – С. 173-176.

26 Гусев П.Ю., Ломаков А.В., Львович Я.Е. Оптимизация управления ресурсным обеспечением в региональной организационной системе на основе предиктивного анализа многолетней статистической информации. Моделирование, оптимизация и информационные технологии. 2024;12(4).

27 Воробьева О.Д. Статистика населения и демография + Приложение: тесты: учебник / О.Д. Воробьева, Т.А. Долбик-Воробей. – М.: КНОРУС, 2019. – 314 с.

28 Рогачева Е.С. Практика применения метода группировки в статистике населения Российской Федерации // Форум молодых ученых. 2017. №11 (15). С. 799-803.

29 Ворошилов Н.В. Особенности и проблемы формирования и использования статистической информации по муниципальным образованиям России // ЭТАП. 2022. №6. С. 89-105.

30 Рязанов И.Е. Сбалансированное управленческое решение: понятие, проблемы оценки в современных условиях // Экономические науки. – М.: ООО «Экономические науки», 2014. – №116. – С. 49-54.

31 Черкасова Т.П. Интегральная оценка эффективности функционирования центра управления компетенциями как инструмента публичного управления / Т.П. Черкасова, О.И. Архипова, Е.А. Аксенов // Journal of Economic Regulation (Вопросы регулирования экономики). 2017. №1. С. 96-103.

- 32 Лобкова Е.В., Петриченко А.С. Управление эффективностью региональной системы здравоохранения. Региональная экономика: Теория и практика. 2018;16(2).
- 33 Я.Е. Львович, И.Я. Львович, О.Н. Чопоров [и др.]. Оптимизация цифрового управления в организационных системах: коллективная монография. Воронеж: ИПЦ «Научная книга»; 2021. 191 с.
- 34 Львович И.Я. Принятие решений на основе оптимизационных моделей и экспертной информации. Воронеж: ИПЦ «Научная книга»; 2023. 232 с.
- 35 Донской В.И. Интеллектуальная оптимизация на основе машинного обучения: современное состояние и перспективы (обзор). Таврический вестник информатики и математики. 2020;46(1). С. 32-63.
- 36 Колесов Д.Н. Совместный учет статистической и экспертной информации при прогнозировании временных рядов экономических показателей* / Д.Н. Колесов, Н.В. Котов, А.С. Федоренко // Вестник Санкт-Петербургского университета. Экономика. 2007. №3. С. 93-101.
- 37 Маршова Т.Н. Принципы формирования статистических данных для анализа и прогноза социально-экономического развития // Экономические и социально-гуманитарные исследования. 2017. №2 (14). С. 25-36.
- 38 Т.Б. Александрова, Д.Е. Калинкин, В.Я. Плещинская, Е.Н. Образцова, Р.М. Тахауов, С.М. Хлынин. Медицинская статистика. Показатели и методы оценки здоровья населения. Томск: СибГМУ; 2011. 126 с.
- 39 Глушанко В.С., Тимофеева А.П., Герберг А.А. Методика изучения уровня, частоты, структуры и динамики заболеваемости и инвалидности. Медико-реабилитационные мероприятия и их составляющие: учеб.-метод. пособие. Витебск: Изд-во ВГМУ; 2016. 177 с.
- 40 Щепин О.П. Региональные аспекты развития здравоохранения. В сборнике: Проблемы социальной гигиены, здравоохранения и истории медицины. 2014;22(5). С. 3-7.

41 Kelleher J.D., Namee B.M., D'Arcy A. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples and Case Studies*. MIT Press, 2020. 856р.

42 Гусев П.Ю. Анализ возможностей предиктивной аналитики при прогнозировании состояния АЭС. В сборнике: известия Тульского государственного университета – Технические науки. 2024; 4.

43 Aloorravi S. *Mastering Time Series Analysis and Forecasting with Python: Bridging Theory and Practice Through Insights, Techniques, and Tools for Effective Time Series Analysis in Python* // Orange Education Pvt Ltd, 2024. 321 p.

44 Van Calster B., Wynants L., Timmerman D., Steyerberg E.W., Collins G.S. Predictive analytics in health care: how can we know it works? *Journal of the American Medical Informatics Association*. 2019;26(12). pp. 1651-1654.

45 Биомедкибернетика [Текст]: монография / [Есауленко И.Э. и др.]; под общ. ред. И.Э. Есауленко. – Воронеж: Истоки, 2014. – 477 с.

46 Садыков С.С., Белякова А.С. Математические модели некоторых сердечно-сосудистых заболеваний, информационные технологии. 2011. №12. С. 59-63.

47 Эпидемиологические методы выявления основных хронических неинфекционных заболеваний и факторов риска при массовых обследованиях населения: методическое пособие / под общей ред. Проф. Бойцова С.М.: Государственный научно-исследовательский центр профилактической медицины. 2015. 96 с.

48 Базалева О.И. Мастерство визуализации данных. СПб: ООО «Диалектика», 2020. 192 С.

49 Паклин Н.Б., Орешков В.И. Визуализация данных / Бизнес-аналитика. От данных к знаниям. – 2-е изд. СПб: Питер, 2013. С. 173-210.

50 Гафанович Е.Я., Ломаков А.В., Львович А.И., Чопоров О.Н. Визуальное и прогностическое моделирование заболеваемости артериальной гипертензией лиц старших возрастных групп и их диспансеризации. Моделирование, оптимизация и информационные технологии. 2024;12(2).

- 51 Крючин О.В. Прогнозирование временных рядов с помощью искусственных нейронных сетей и регрессивных моделей на примере прогнозирования котировок валютных пар. Электронный научный журнал «Исследования в России». 2010. №30. С. 354-362.
- 52 Слободенюк А.В., Косова А.А., Ан Р.Н. Эпидемиологический анализ: учебное пособие. Екатеринбург: изд-во ГБОУ ВПО УГМУ минздрава России, 2015. 36 с.
- 53 Драпкина О.М., Дроздова Л.Ю., Калинина А.М. и др. Организация проведения профилактического медицинского осмотра и диспансеризации определенных групп взрослого населения: методические рекомендации / под ред. О.М. Драпкиной. М.: ФГБУ «НМНЦТПМ» Минздрава России, 2020. 232 с.
- 54 Горшкова Л.В. Проблемы оценки эффективности затрат на здравоохранение. Сервис в России и за рубежом. 2017. №6 (76). С. 137-151.
- 55 Нильсен Э. Практический анализ временных рядов. Прогнозирование со статистикой и машинное обучение: Пер. с англ. – СПб.: ООО «Диалектика», 2021. – 544 с.
- 56 Замятин А.В. Интеллектуальный анализ данных: учеб. пособие. – Томск: Издательский Дом Томского государственного университета, 2016. – 120 с.
- 57 Семенова А.А. Путь машинного обучения: междисциплинарный анализ развития технологий // Проблемы деятельности ученого и научных коллективов. 2024. №10 (40). С. 98-110.
- 58 Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. Пер. с англ. – СПб.: ООО "Альфа-книга": 2018. – 688 с.
- 59 Чучуева И.А. Модель прогнозирования временных рядов по выборке максимального подобия, диссертация ... канд. тех. наук / Московский государственный технический университет им. Н.Э. Баумана. Москва, 2012.
- 60 Морозова В.И. Прогнозирование методом машинного обучения / В.И. Морозова, Д.И. Логунова // Молодой ученый. – 2022. – № 21 (416). – С. 202-204.

- 61 Ал-Зубаиди Азхар Кхудхаир Аббас Применение машинного обучения в диагностике туберкулеза: систематический обзор литературы // Медицина. Социология. Философия. Прикладные исследования. 2023. №5. – С. 8-14.
- 62 Шадеркин И.А. Анализ влияния факторов окружающей среды на здоровье человека с применением цифровых решений. Мнение эксперта / И.А. Шадеркин, А.П. Дьяченко, Е.А. Чулюкова, Т.В. Пшеничный, Л.В. Ковека, В.Е. Храмцова, Н.С. Гугняевых, А.И. Кузьмина // Журнал телемедицины и электронного здравоохранения. 2023. №4. – С. 36-44.
- 63 Герасимова А.Е. Анализ налоговой чувствительности отдельных отраслей экономики // Налоги и налогообложение. 2024. №4. – С. 64-79.
- 64 Тяжельникова Ю.А. Использование нейронных сетей для прогнозирования инфляции / Ю.А. Тяжельникова, Ю.В. Шершон, В.В. Драгуленко // Региональная и отраслевая экономика. 2023. №4. – С. 172-177.
- 65 Казаковцева М.В. Прогнозирование котировок финансовых инструментов с помощью нейронных сетей / М.В. Казаковцева, Е.В. Конакова // Вестник Марийского государственного университета. Серия «Сельскохозяйственные науки. Экономические науки». 2023. №4. – С. 433-442.
- 66 Царев С.А. Использование технологий искусственного интеллекта (ml и нейросетей) для прогнозирования уровня смертности пациентов, страдающих наркологическими заболеваниями / С.А. Царев, А.В. Щербань, А.С. Бенян, И.И. Сиротко, А.А. Савинцев // Современные проблемы здравоохранения и медицинской статистики. 2024. №3. – С. 286-301.
- 67 Recursive multi-step forecasting [Электронный ресурс]. – URL: https://skforecast.org/0.15.1/user_guides/autoregressive-forecaster#create-and-train-forecaster (Дата обращения: 02.02.2025).
- 68 Гафанович Е.Я. Формирование ансамбля прогностических моделей на основе предиктивного анализа многолетней медико-статистической информации / Е.Я. Гафанович, А.В. Ломаков, М.М. Мурадханов // Оптимизация и моделирование в автоматизированных системах: труды Международной молодежной научной школы / отв. ред. Я.Е. Львович. – Воронеж: ВГТУ, 2025. – С. 153-155.

69 Боброва О.П., Шнайдер Н.А., Петрова М.М., Зобова С.Н. и др. Прогнозирование фентанил-ассоциированной нейротоксичности у больных с раком поджелудочной железы с помощью клинико-генетической модели // ЭиКГ. 2021. №3 (187). С. 136-145.

70 Перова М.Д., Самохвалова Д.Д., Халафян А.А., Акиньшина В.А. Определение относительного риска прогрессирования пародонтита с помощью нейросетевого моделирования: когортное ретроспективное исследование // Кубанский научный медицинский вестник. 2022. №5. С. 44-62.

71 Ломаков А.В. Интеллектуализация управления в региональной организационной системе здравоохранения при распределении ресурсного обеспечения / А.В. Ломаков, П.Ю. Гусев // Интеллектуальные информационные системы: труды Международной научно-практической конференции / отв. ред. Я.Е. Львович – Воронеж, ВГТУ, 2025.

72 Бринк Х., Ричардс Дж., Феверолф М. Машинное обучение. – СПб.: Питер, 2017. – 336 с.

73 Демидова Л.А., Соколова Ю.С. Интеллектуальный анализ данных в пакете STATISTICA: методические указания – Рязань: Рязан. гос. радиотехн. ун-т, 2018. – 32 с.

74 Шунина Ю.С. Критерии качества работы классификаторов [Текст] / Ю.С. Шунина, В.А. Алексеева, В.Н. Клячкин // Вестник УлГТУ. – 2015. – №2. – С. 67-70.

75 Уэс Маккинли Python и анализ данных / Пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 482 С.

76 Scikit-Learn [Электронный ресурс]. Scikit-Learn Machine Learning in Python. URL: <https://scikit-learn.org/> (Дата обращения: 10.09.2024).

77 Орлов А.И. О развитии теории принятия решений и экспертных оценок // Научный журнал КубГАУ. 2021. №167. – С. 177-198.

78 Львович Я.Е. Многоальтернативная оптимизация: теория и приложения. Воронеж: Издательский дом «Кварт»; 2006. 426 с.

79 Львович И.Я. Информационные технологии моделирования и оптимизации: краткая теория и приложения. / И.Я. Львович, Я.Е. Львович, В.Н. Фролов // Воронеж: ИПЦ "Начная книга", 2016. – 444 с.

- 80 Подиновский В.В. Парето-оптимальное решение многоокритериальных задач / В.В. Подиновский, В.Д. Ногин. – Москва: Наука, 1986. 256 с.
- 81 Львович Я.Е. Теоретические основы конструирования, технологии и надежности РЭА: Учебное пособие для вузов / Я.Е. Львович, В.Н. Фролов. – Москва: Радио и связь, 1986. – 192 с.
- 82 Паклин Н. Б., Орешков В. И. Визуализация данных // Бизнес-аналитика. От данных к знаниям. – 2-е изд. – СПб.: Питер, 2013. – С. 173-210.
- 83 Красочкин С.Г. Изображения и визуализация данных в Python // Научный журнал. 2022. №2 (64). – С. 5-8.
- 84 Вандер Плас Дж. Python для сложных задач: наука о данных. 2-е межд. изд. СПб.: Издательский дом «Питер»; 2024. 595 с.
- 85 Рашка С. Python и машинное обучение. М.: ДМК Пресс; 2017. 418 с.
- 86 Мюллер А., Гвидо С. Машинное обучение с помощью Python. Руководство для специалистов по работе с данными. М.: Издательский дом "Вильямс"; 2022. 480 с.
- 87 Галлатин К. Машинное обучение с использованием Python. Сборник рецептов: Пер. с англ. / К. Галлатин, К. Элбон. - 2-е изд., перераб. и доп. – Астана: АЛИСТ, 2024. – 448 с.
- 88 Келлехер Д. Дж. Основы машинного обучения для аналитического прогнозирования: алгоритмы, рабочие примеры и тематические исследования / Дж. Д. Келлехер, Б. Мак-Нейми, А. д'Арси // М.: Издательский дом "Вильямс"; 2019. 656 с.
- 89 Анафиев А.С. Обзор подходов к решению задачи оптимизации гиперпараметров для алгоритмов машинного обучения / А.С. Анафиев, А.С. Карюк // ТВИМ. 2022. №2 (55). С. 30-37.
- 90 Прохоренок Н.А. Python 3 и PyQt5. Разработка приложений / Н.А. Прохоренок, В.А. Дронов. – СПб.: БХВ-Петербург, 2016. – 832 С.
- 91 Чан У. Python: создание приложений. Библиотека профессионала, 3-е изд. Пер. с англ. – М.: ООО "И.Д. Вильяме", 2016. – 816 с.
- 92 Арнагельдыева А., Корреляционная связь // Наука и мировоззрение. – 2025. – 47(1). С. 38-43.

ПРИЛОЖЕНИЕ А. Акты внедрения



ООО «Актив Компьютерс», ИНН 7705711670,
КПП 772001001;
111123, г. Москва, шоссе Энтузиастов, д. 31, стр. 39,
эт. 3, оф. 2,3;
+7 (495) 950-57-99; pay@acomps.ru; www.acomps.ru

«17» август 2023 г.

АКТ

внедрения результатов проектной работы

Ломакова Андрея Владимировича

Настоящим подтверждаю, что результаты проектной работы, выполненной в Воронежском институте высоких технологий – автономной некоммерческой образовательной организации высшего образования под руководством к.т.н., доц. Зелениной А.Н. обучающимся Ломаковым А.В. (аспирант гр. А-201) на тему ИНТЕЛЛЕКТУАЛЬНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ БИНАРНОЙ КЛАССИФИКАЦИИ ДАННЫХ внедрены в производственную деятельность ООО «Актив Компьютерс».

Аспирантом было разработано программное обеспечение, способное проводить обучение классификационных моделей на основе внутренних наборов данных, составленных в организации, а также осуществлять прогнозирование в коммерческих целях компании на основе обученных классификаторов.

Разработка внедрена в модули обработки потоков данных аудита и мониторинга аппаратных ресурсов, разрабатываемого программного продукта U-Connect. Результаты проектной работы позволили значительно сократить время обработки в высоконагруженных системах.

Директор по развитию, к.т.н.





Председатель НТС

И.Я. Львович

«10» апреля 2025г.

АКТ

о внедрении в учебный процесс

Воронежского института высоких технологий – автономной некоммерческой
образовательной организации высшего образования
результатов научных исследований

Тема научного исследования: «Предиктивный анализ на основе машинного обучения».

Автор: Ломаков Андрей Владимирович.

Научный руководитель: д.т.н., проф. Львович Я. Е.

В период с 01.09.2024 и по наст. вр. внедрены в учебный процесс ВИВТ-АНОО ВО по направлениям подготовки 09.03.01. «Информатика и вычислительная техника» (профиль подготовки «Автоматизированные системы обработки информации и управления») и 09.04.01 «Информатика и вычислительная техника» (Программа «Киберфизические системы и искусственный интеллект» профиль «Облачная и сетевая инфраструктура систем искусственного интеллекта») на основании решения Научно-технического совета ВИВТ- АНОО ВО от «09» 04 2025 г. протокол № 8.

1. **Вид результатов, внедренных в учебный процесс:** совокупность знаний и представлений по теме научного исследования (разработка лекций на основе результатов исследования; составление вопросов для устного опроса студентов с использованием результатов исследования; составление задания лабораторной работы для студентов,

основанных на результатах исследования; разработка вопросов для экзамена (зачета с оценкой), которые будут включать в себя знания студента по результатам исследования).

2. **Область применения:** учебно-методический комплекс по дисциплинам «Теория оптимизации и принятия решений в автоматизированных системах», «Технологии машинного обучения и искусственного интеллекта», «Технологии цифровой трансформации» а также при выполнении курсовых и дипломных работ.

3. **Форма внедрения:** описание разработанных в научном исследовании методов и процессов прогнозирования в предиктивной аналитике на основе выбора модели и типа нейросетевого обучения с учетом подбора подходящего алгоритма в зависимости от поставленной задачи и типа полученной информации.

Актуальность данной темы с точки зрения учебного процесса заключается в возможности применения полученных результатов для совершенствования методик преподавания дисциплин, связанных машинным обучением и искусственным интеллектом в рамках предиктивного анализа многолетней статистической информации. Внедрение результатов научных исследований позволит обучающимся получить навыки прогнозирования временных рядов с использованием методов машинного обучения, что дает возможность использовать полученные знания в будущей профессиональной деятельности.

Целью исследования явилось повышение эффективности управления в региональной организационной системе на основе интеграции результатов предиктивного анализа многолетней статистической информации в процесс принятия управленческих решений.

Задачи исследования состояли в:

- исследовании особенностей использования многолетней статистической информации при управлении в региональной организационной системе;
- изучении структуры предиктивного анализа и этапов процесса прогнозирования;
- исследовании методов прогнозирования временных рядов с использованием методов машинного обучения;
- внедрении в методику обучения результатов собственных научных исследований с учетом современных тенденций развития информационных технологий;
- подготовка методических материалов по результатам собственных научных исследований с целью применения в учебном процессе.

4. Эффект от внедрения: **повышение** качества образовательного процесса на основе применения теоретических и практических знаний, полученных в ходе научного исследования с учетом современных методик нейросетевого моделирования.

Председатель ИТС  д.т.н., профессор Львович И.Я.

Секретарь ИТС  Линкина А.В.

УТВЕРЖДАЮ
И.о. Начальника ФГБУЗ МСЧ 97
Федерального медико-биологического агентства России

Данилов Д.А.
«10» 12 2025г.



АКТ
о внедрении результатов диссертационного исследования
Ломакова Андрея Владимировича «Интеллектуализация процессов
управления в рамках региональной отраслевой организационной
системы на основе средств предиктивного анализа
ретроспективной информации»

Разработанные Ломаковым А.В. модели и процедуры для интеллектуальной поддержки принятия управленческих решений при распределении ресурсного обеспечения, позволяют повысить эффективность процесса управления в практике прогнозирования объемов диспансеризации различных групп взрослого населения. Предложенные автором алгоритм предиктивного анализа многолетней статистики и оптимизационные модели обеспечивают сокращение времени на выработку управленческих решений при распределении планового объема ресурса.

Разработанные автором программные модули позволяют получить прогнозы по заболеваемости на будущие временные периоды, используя более широкий охват ретроспективных данных. Моделирование управленческих и скорректированных решений с использованием средств интеллектуализации принятия решений позволяют сбалансировать административные управленческие действия с учетом множества показателей, фиксируемых собираемой статистикой.

Результаты диссертационного исследования Ломакова Андрея Владимировича «Интеллектуализация процессов управления в рамках региональной отраслевой организационной системы на основе средств предиктивного анализа ретроспективной информации» планируются к внедрению в деятельность организации в 2026 году.

Комиссия:

Начальник отдела АСУ и обработки медико-статистической информации

Романов Г.Е. Ф.И.О.

Заведующий поликлиникой

Мирошникова И.Р. Ф.И.О.

ПРИЛОЖЕНИЕ Б. Свидетельства о регистрации программ для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2021616488

«Интеллектуальное программное обеспечение бинарной классификации данных с использованием алгоритма k -ближайших соседей» (Intelligent binary data classification software using k -nearest neighbors algorithm)

Правообладатель: **Федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (RU)**

Автор(ы): **Ломаков Андрей Владимирович (RU)**

Заявка № **2021615463**

Дата поступления **13 апреля 2021 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **22 апреля 2021 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ибрагимов



РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023688819

Программное обеспечение классификации данных с
использованием базовых и ансамблевых методов

Правообладатель: *Воронежский институт высоких
технологий - автономная некоммерческая
образовательная организация высшего образования (RU)*

Авторы: *Ломаков Андрей Владимирович (RU), Львович Яков
Евсеевич (RU), Преображенский Андрей Петрович (RU)*

Заявка № 2023686895

Дата поступления 06 декабря 2023 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 25 декабря 2023 г.

Руководитель Федеральной службы
по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 429b6a0fe3553164ba956f83b73b4aa7
Владелец Зубов Юрий Сергеевич
Действителен с 10.05.2023 по 02.08.2024

Ю.С. Зубов



**ПРИЛОЖЕНИЕ В. Сравнение результатов работы методов
прогнозирования**

Таблица В.1 – Сравнение результатов работы методов прогнозирования для данных по первичной заболеваемости (болезни, характеризующиеся повышенным кровяным давлением)

Группа выборки	Модель	RMSE	MAE	R2
Группа 1	ARIMA	905.71	901.2	-732.14
	HL	62.64	62.64	-2.51
	HW	1300.82	1300.27	-1511.31
	DT	97.91	84.0	-7.57
	RF	129.08	112.94	-13.89
	MLP	517.61	517.22	-238.45
Группа 2	ARIMA	736.62	576.03	-1.57
	HL	953.89	874.66	-3.32
	HW	914.08	832.29	-2.96
	DT	2148.94	1928.14	-20.91
	RF	1453.02	1304.76	-9.02
	MLP	860.02	774.41	-2.51
Группа 3	ARIMA	946.55	945.89	0.05
	HL	870.09	804.28	0.19
	HW	945.77	883.54	0.05
	DT	2104.24	1715.65	-3.71
	RF	1994.35	1606.39	-3.23
	MLP	1715.65	1413.28	-2.13
Воронеж	ARIMA	1169.66	1052.87	-2.56
	HL	1470.73	1430.97	-4.63
	HW	1046.43	836.35	-1.85
	DT	1538.55	1287.5	-5.16
	RF	1820.09	1569.69	-7.63
	MLP	1152.75	896.1	-2.46
Область	ARIMA	776.98	710.62	-0.31
	HL	1015.11	882.24	-1.24
	HW	1041.14	898.6	-1.35
	DT	2242.01	1951.95	-9.92
	RF	1768.41	1484.39	-5.79
	MLP	1018.01	761.23	-1.25

Таблица В.2 – Сравнение результатов работы методов прогнозирования для данных по числу больных с впервые в жизни установленным диагнозом, взятых под диспансерное наблюдение (болезни, характеризующиеся повышенным кровяным давлением)

Группа выборки	Модель	RMSE	MAE	R2
Группа 1	ARIMA	365.18	327.48	-118.19
	HL	5124.94	4812.31	-23472.89
	HW	724.64	694.21	-468.31
	DT	123.03	102.2	-12.53
	RF	129.86	113.12	-14.07
	MLP	72.66	66.04	-3.72
Группа 2	ARIMA	598.7	448.87	-0.78
	HL	2887.61	2854.09	-40.38
	HW	813.27	734.01	-2.28
	DT	1354.24	1268.44	-8.1
	RF	1191.25	1045.51	-6.04
	MLP	715.52	639.49	-1.54
Группа 3	ARIMA	1589.68	1220.07	-1.72
	HL	3219.87	3209.92	-10.16
	HW	1850.17	1315.52	-2.69
	DT	2061.83	1724.8	-3.58
	RF	1990.11	1614.27	-3.26
	MLP	1421.78	1013.65	-1.18
Воронеж	ARIMA	920.22	660.87	-0.99
	HL	1425.37	1382.58	-3.77
	HW	1143.5	974.83	-2.07
	DT	1542.79	1279.75	-4.59
	RF	1788.4	1521.78	-6.51
	MLP	1418.73	1061.39	-3.73
Область	ARIMA	1259.72	1066.48	-2.37
	HL	2225.55	2225.13	-9.53
	HW	816.05	805.59	-0.42
	DT	1738.03	1541.85	-5.42
	RF	1638.5	1389.63	-4.7
	MLP	1151.82	851.55	-1.82

Таблица В.3 – Сравнение результатов работы методов прогнозирования для данных по числу больных с впервые в жизни установленным диагнозом, выявленных при профосмотре (болезни, характеризующиеся повышенным кровяным давлением)

Группа выборки	Модель	RMSE	MAE	R2
Группа 1	ARIMA	0.0	0.0	1.0
	HL	0.0	0.0	1.0
	HW	0.0	0.0	1.0
	DT	0.0	0.0	1.0
	RF	0.0	0.0	1.0
	MLP	0.0	0.0	1.0
Группа 2	ARIMA	8.65	7.62	-330.89
	HL	18.62	15.93	-1536.19
	HW	8.98	8.41	-356.67
	DT	16.73	15.77	-1239.69
	RF	14.61	13.79	-944.89
	MLP	20.83	19.47	-1921.44
Группа 3	ARIMA	37.41	37.24	-477.53
	HL	19.39	14.07	-127.52
	HW	15.49	12.82	-81.02
	DT	50.26	47.32	-862.78
	RF	37.09	35.18	-469.47
	MLP	12.21	10.86	-49.99
Воронеж	ARIMA	17.56	14.38	-0.51
	HL	78.15	77.63	-28.87
	HW	16.45	14.56	-0.32
	DT	57.01	56.7	-14.89
	RF	32.85	31.98	-4.28
	MLP	10.6	9.16	0.45
Область	ARIMA	16.59	15.28	-5.72
	HL	18.1	17.14	-7.0
	HW	12.03	12.03	-2.53
	DT	32.09	31.9	-24.14
	RF	28.51	26.43	-18.84
	MLP	13.14	10.21	-3.22

Таблица В.4 – Сравнение результатов работы методов прогнозирования для данных по числу больных с впервые в жизни установленным диагнозом, выявленных при диспансеризации (болезни, характеризующиеся повышенным кровяным давлением)

Группа выборки	Модель	RMSE	MAE	R2
Группа 1	ARIMA	716.38	716.02	-27754.89
	HL	3527.92	3277.61	-673129.92
	HW	386.51	367.5	-8078.61
	DT	5.52	4.1	-0.65
	RF	112.38	109.6	-682.03
	MLP	21.75	21.53	-24.58
Группа 2	ARIMA	551.38	526.34	-8.33
	HL	339.6	330.78	-2.54
	HW	461.73	430.48	-5.55
	DT	877.8	868.0	-22.66
	RF	687.95	629.62	-13.53
	MLP	637.61	575.55	-11.48
Группа 3	ARIMA	1477.59	1128.98	-0.72
	HL	1344.28	1093.58	-0.42
	HW	1515.9	1134.78	-0.81
	DT	1960.32	1506.77	-2.02
	RF	1802.44	1349.38	-1.56
	MLP	1760.49	1271.71	-1.44
Воронеж	ARIMA	654.08	620.89	-0.11
	HL	666.49	599.48	-0.15
	HW	710.86	531.26	-0.31
	DT	1045.81	795.45	-1.84
	RF	1488.18	1245.14	-4.75
	MLP	1387.89	1018.85	-4.0
Область	ARIMA	587.33	482.45	-0.27
	HL	606.66	606.13	-0.36
	HW	596.45	445.56	-0.31
	DT	1702.86	1480.95	-9.71
	RF	1257.28	1054.51	-4.84
	MLP	936.48	691.78	-2.24