

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.03 Прикладная информатика
Отделение информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

| Тема работы |
|--|
| Информационная технология анализа результатов лечения детей с различными формами ожирения |

УДК 004:613.25-053.3-08

Студент

| Группа | ФИО | Подпись | Дата |
|--------|--------------------------|---------|------|
| 8КМБ1 | Лызин Иван Александрович | | |

Руководитель

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|---------------------|--------------------------------|---------------------------|---------|------|
| Доцент ОИТ ИШИТР | Марухина Ольга Владимировна | К.Т.Н. | | |

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|-------------|-----------------------------------|---------------------------|---------|------|
| Доцент ОСГН | Старикова Екатерина Васильевна | к.ф.н. | | |

По разделу «Социальная ответственность»

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|------------------|------------------------|---------------------------|---------|------|
| Ассистент ОКД | Авдеева Ирина Ивановна | - | | |

ДОПУСТИТЬ К ЗАЩИТЕ:

| Руководитель ООП 09.04.03 Прикладная информатика | ФИО | Ученая степень, звание | Подпись | Дата |
|---|--------------------------------|---------------------------|---------|------|
| Доцент ОИТ ИШИТР | Марухина Ольга Владимировна | К.Т.Н. | | |

Томск – 2018 г.

**Планируемые результаты обучения по программе
09.04.03 – Прикладная информатика**

| Код | Результат обучения |
|--|--|
| P1 | Применяет базовые и специальные знания в области современных информационно-коммуникационных технологий для решения междисциплинарных инженерных задач. |
| P2 | Проводит теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретацию полученных данных в области информатизации и автоматизации прикладных процессов и создания, внедрения, эксплуатации и управления информационными системами в прикладных областях |
| P3 | Внедряет, сопровождает и эксплуатирует современные информационные системы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья и безопасности труда, выполнять требования по защите окружающей среды |
| P4 | Активно владеет иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности. |
| P5 | Владеет и применяет методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе глобальных компьютерных сетей. |
| P6 | Эффективно работает индивидуально, в качестве члена и руководителя группы, состоящей из специалистов различных направлений и квалификаций, демонстрирует ответственность за результаты работы и готовность следовать корпоративной культуре организации |
| P7 | Самостоятельно учится и непрерывно повышает квалификацию в течение всего периода профессиональной деятельности |
| Профиль «Системы корпоративного управления» | |
| P8 | Применяет глубокие профессиональные знания основ построения информационных технологий и систем, достаточные для решения научных и профессиональных задач производства. Знает современные проблемы и методы прикладной информатики и научно-технического развития информационных технологий. |
| P9 | Ставит и решает задачи комплексного анализа, связанные с информатизацией и автоматизацией прикладных процессов; созданием, внедрением, эксплуатацией и управлением информационными системами в прикладных областях, с использованием базовых и специальных знаний, современных аналитических методов и моделей |
| P10 | Способен организовывать работы по моделированию прикладных ИС и реинжинирингу прикладных и информационных процессов предприятия и организации. Способен управлять проектами по информатизации прикладных задач и созданию ИС предприятий и организаций. |

Министерство образования и науки Российской Федерации
 федеральное государственное автономное образовательное учреждение
 высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
 ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.03. Прикладная информатика
 Отделение информационных технологий

УТВЕРЖДАЮ:

Руководитель ООП 09.04.03. Прикладная информатика

 (Подпись) (Дата) Марухина О.В.
 (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

| |
|--------------------------|
| магистерской диссертации |
|--------------------------|

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

| Группа | ФИО |
|--------|-----------------------------|
| 8КМ61 | Лызину Ивану Александровичу |

Тема работы:

| | |
|---|-----------------------|
| Информационная технология анализа результатов лечения с различными формами ожирения | |
| Утверждена приказом директора (дата, номер) | 20.04.2018 г., 2796/с |

| | |
|--|--|
| Срок сдачи студентом выполненной работы: | |
|--|--|

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

| | |
|--|---|
| <p>Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p> | <p>Клинико-лабораторные показатели НИИ курортологии и физиотерапии</p> |
| <p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p> | <p>Постановка задачи Анализ предметной области Выбор решения задачи Решение поставленной задачи Описание полученных результатов</p> |
| <p>Перечень графического материала <i>(с точным указанием обязательных чертежей)</i></p> | <p>Диаграммы размахов Графики плотности распределения временных данных Визуальные образы на основе лиц Чернова</p> |

| Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i> | |
|--|--------------------------------|
| Раздел | Консультант |
| Разделы 1-3 | Марухина Ольга Владимировна |
| Раздел 4 | Старикова Екатерина Васильевна |
| Раздел 5 | Авдеева Ирина Ивановна |
| Названия разделов, которые должны быть написаны на русском и иностранном языках: | |
| Введение | |
| Обзор литературы | |
| Глава 1 Постановка задачи | |
| Глава 2 Выбор решения задачи | |

| | |
|---|--|
| Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику | |
|---|--|

Задание выдал руководитель:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|---------------------|-----------------------------|-------------------------------|----------------|-------------|
| Доцент ОИТ ИШИТР | Марухина Ольга Владимировна | К. Т. Н. | | |

Задание принял к исполнению студент:

| Группа | ФИО | Подпись | Дата |
|---------------|--------------------------|----------------|-------------|
| 8КМб1 | Лызин Иван Александрович | | |

Министерство образования и науки Российской Федерации
 Федеральное государственное автономное образовательное учреждение
 высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
 ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники

Специальность 09.04.03. Прикладная информатика

Уровень образования Магистратура

Отделение Информационных технологий

Период выполнения Весенний семестр 2018 учебного года

Форма представления работы:

магистерская диссертация

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
 выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:

| Дата контроля | Название раздела (модуля) / вид работы (исследования) | Максимальный балл раздела (модуля) |
|---------------|--|------------------------------------|
| 01.02.2018 | Получение задания на ВКР | |
| 01.03.2018 | Получение задания по финансовому менеджменту | |
| 01.03.2018 | Получение задания по социальной ответственности | |
| 19.03.2018 | Глава 1. Постановка задачи | |
| 02.04.2018 | Глава 2. Изучение и выбор технологии, инструментов и методов | |
| 23.04.2018 | Глава 3. Решение поставленной задачи и описание результатов | |
| 30.04.2018 | Глава 4. Финансовый менеджмент | |
| 07.04.2018 | Глава 5. Социальная ответственность | |
| 20.05.2018 | Проверка оформления. Заключение. | |
| 29.05.2018 | Сдача выполненной работы. | |

Составил преподаватель:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|------------|-----------------------------|------------------------|---------|------|
| Доцент ОИТ | Марухина Ольга Владимировна | К. Т. Н. | | |

СОГЛАСОВАНО:

| Руководитель ООП 09.04.03 Прикладная информатика | ФИО | Ученая степень, звание | Подпись | Дата |
|--|-----------------------------|------------------------|---------|------|
| Доцент ОИТ ИШИТР | Марухина Ольга Владимировна | К. Т. Н. | | |

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

| | |
|---------------|-----------------------------|
| Группа | ФИО |
| 8КМ61 | Лызину Ивану Александровичу |

| | | | |
|----------------------------|--|----------------------------------|---------------------------------|
| Школа | Инженерная школа информационных технологий и робототехники | Отделение | Информационных технологий |
| Уровень образования | Магистратура | Направление/специальность | 09.04.03 Прикладная информатика |

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

| | |
|---|---|
| 1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i> | <i>Работа с информацией о состоянии здоровья человека, а именно общая информация о человеке, показатели его анализов в разные периоды времени. Разработанная информационная технология позволяет восстанавливать пропуски в исходных наборах данных и комплексно оценивать результаты лечения посредством визуальных образов.</i> |
| 2. <i>Нормы и нормативы расходования ресурсов</i> | |
| 3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i> | |

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

| | |
|--|--|
| 1. <i>Оценка коммерческого и инновационного потенциала НТИ</i> | <i>Оценка конкурентоспособности, рассмотрение альтернатив проведения НИ, SWOT анализ, QuaD технология.</i> |
| 2. <i>Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок</i> | <i>Планирование этапов разработки информационной технологии, определение трудоемкости, построение диаграммы Ганта.</i> |
| 3. <i>Определение ресурсной, финансовой, экономической эффективности</i> | <i>Сравнительный анализ интегральных показателей эффективности, формирование бюджета НТИ</i> |

Перечень графического материала (с точным указанием обязательных чертежей):

| |
|--|
| 1. Оценка конкурентоспособности технических решений |
| 2. SWOT анализ, QuaD технология |
| 3. График проведения и бюджет НИ |
| 4. Оценка ресурсной, финансовой и экономической эффективности НИ |

| | |
|---|--|
| Дата выдачи задания для раздела по линейному графику | |
|---|--|

Задание выдал консультант:

| | | | | |
|------------------|--------------------------------|-------------------------------|----------------|-------------|
| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
| Доцент ОСГН | Старикова Екатерина Васильевна | к. ф. н | | |

Задание принял к исполнению студент:

| | | | |
|---------------|--------------------------|----------------|-------------|
| Группа | ФИО | Подпись | Дата |
| 8КМ61 | Лызин Иван Александрович | | |

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

| | |
|---------------|-----------------------------|
| Группа | ФИО |
| 8КМ61 | Лызину Ивану Александровичу |

| | | | |
|----------------------------|--|----------------------------------|---------------------------------|
| Школа | Инженерная школа информационных технологий и робототехники | Отделение | Информационных технологий |
| Уровень образования | Магистратура | Направление/специальность | 09.04.03 Прикладная информатика |

Исходные данные к разделу «Социальная ответственность»:

| | |
|--|--|
| <p>1. Описание рабочего места (рабочей зоны, технологического процесса, механического оборудования) на предмет возникновения:</p> <ul style="list-style-type: none"> – вредных проявлений факторов производственной среды (метеоусловия, вредные вещества, освещение, шумы, вибрации, электромагнитные поля, ионизирующие излучения) – опасных проявлений факторов производственной среды (механической природы, термического характера, электрической, пожарной и взрывной природы) – негативного воздействия на окружающую природную среду (атмосферу, гидросферу, литосферу) – чрезвычайных ситуаций (техногенного, стихийного, экологического и социального характера) | <p>Рабочее место – кабинет №204 инженерной школы информационных технологий и робототехники Национального Исследовательского Томского Политехнического Университета.</p> <p>Данный кабинет представляет из себя помещение площадью 36 м² (6м*6м). Стены и потолок исполнены в светлых тонах. Пол бетонный, покрытый линолеумом светлого оттенка. В помещении имеется два окна (размер 1х1,35 м). Освещение естественное только в светлое время суток, по большей части в теплое время года. В остальные времена года превалирует общее равномерное искусственное освещение. Рабочим местом научных работников являются оборудованный компьютерный стол и офисный стул.</p> |
|--|--|

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

| | |
|---|--|
| <p>1. Производственная безопасность</p> <p>1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности:</p> <ul style="list-style-type: none"> – физико-химическая природа вредности, её связь с разрабатываемой темой; – действие фактора на организм человека; – приведение допустимых норм с необходимой размерностью (со ссылкой на соответствующий нормативно-технический документ); – предлагаемые средства защиты; – (сначала коллективной защиты, затем – индивидуальные защитные средства). <p>1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности:</p> <ul style="list-style-type: none"> – механические опасности (источники, средства защиты); – термические опасности (источники, средства защиты); – электробезопасность (в т.ч. статическое электричество, молниезащита – источники, средства защиты); – пожаровзрывобезопасность (причины, профилактические мероприятия, первичные средства пожаротушения). | <p>Анализ выявленных вредных факторов:</p> <ul style="list-style-type: none"> – Недостаточная освещенность рабочей зоны; – Электромагнитные излучения; – Микроклимат; – Шум. <p>Анализ психофизических факторов:</p> <ul style="list-style-type: none"> – Монотонность труда; – Статические нагрузки; – Перенапряжение зрения; – Умственное перенапряжение. <p>Анализ выявленных опасных факторов:</p> <ul style="list-style-type: none"> – Статическое электричество; – Короткое замыкание; – Электрический ток – Механические повреждения – Термический |
| <p>2. Экологическая безопасность:</p> <ul style="list-style-type: none"> – защита селитебной зоны | <p>Анализ негативного воздействия на окружающую природную среду: переработку</p> |

| | |
|--|--|
| <ul style="list-style-type: none"> – анализ воздействия объекта на атмосферу (выбросы); – анализ воздействия объекта на гидросферу (сбросы); – анализ воздействия объекта на литосферу (отходы); – разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды. | <p>бумажных отходов, утилизация люминесцентных ламп, компьютеров и другой оргтехники.</p> |
| <p>3. Безопасность в чрезвычайных ситуациях:</p> <ul style="list-style-type: none"> – перечень возможных ЧС при разработке и эксплуатации проектируемого решения; – выбор наиболее типичной ЧС; – разработка превентивных мер по предупреждению ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий. | <p>Возможные чрезвычайные ситуации на объекте:</p> <ul style="list-style-type: none"> – Пожар; – Пожароопасность; – Землетрясение. |
| <p>4. Правовые и организационные вопросы обеспечения безопасности:</p> <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. | <ul style="list-style-type: none"> – Организация рабочих мест с электронно-вычислительными машинами регулируется СанПиНом 2.2.2/2.4.1340-03. – Рабочее место при выполнении работ сидя регулируется Гост 12.2.032-78 ССБТ. – Закон Томской области от 9 июля 2003 года № 83-ОЗ Об охране труда в Томской области (с изменениями на 4 июля 2014 года). – "Трудовой кодекс Российской Федерации" от 30.12.2001 N 197-ФЗ (ред. от 05.02.2018) |

| | |
|---|------------|
| Дата выдачи задания для раздела по линейному графику | 01.03.2018 |
|---|------------|

Задание выдал консультант:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|---------------|------------------------|------------------------|---------|------|
| Ассистент ОКД | Авдеева Ирина Ивановна | | | |

Задание принял к исполнению студент:

| Группа | ФИО | Подпись | Дата |
|--------|--------------------------|---------|------|
| 8КМ61 | Лызин Иван Александрович | | |

РЕФЕРАТ

Выпускная квалификационная работа содержит 123 страницы, 15 рисунков, 26 таблиц, 53 источника, 1 приложение.

Ключевые слова: Data Mining, многомерные данные, восстановление данных, анализ, информационная технология, визуальные образы.

Объект исследования: многомерные данные (клинико-лабораторные показатели лечение детей и подростков с проблемами ожирения).

Предмет исследования – обработка многомерных данных с использованием пакетов языка R.

Целью работы является создание информационной технологии оценки результатов лечения пациентов, страдающих ожирением разной степени, проводимого по пяти различным траекториям.

Научной новизной исследования является алгоритм оценки результатов лечения пациентов, заложенный в основу информационной технологии.

Метод исследования – теоретический, практический, получение информации о предметной области, её исследование и анализ.

В процессе исследования, проведено визуальное исследование структуры пропущенных данных клинико-лабораторных показателей, осуществлено восстановление пропущенных данных с использованием метода множественного восстановления на основе полученных результатов построены визуальные образы с применением такого типа пиктографиков, как лица Чернова.

В результате разработана информационная технология, реализующая основные функции: анализ выбросов, анализ пропусков, восстановление пропусков, построение визуальных образов.

Выпускная квалификационная работа выполнена в текстовом редакторе Microsoft Word 2016 и представлена на диске (в конверте на обороте обложки).

Оглавление

| | |
|--|----|
| Введение..... | 12 |
| Обзор литературы..... | 14 |
| Глава 1 Постановка задачи..... | 20 |
| 1.1 Информационные технологии анализа данных | 20 |
| 1.2 Описание предметной области | 22 |
| Глава 2 Выбор методов решения задачи | 27 |
| 2.1 Выбор технологии и описание методов решения задачи | 27 |
| 2.1.1 Технология Data Mining | 27 |
| 2.1.2 Анализ выбросов..... | 28 |
| 2.1.3 Анализ пропусков | 28 |
| 2.1.4 Метод множественного восстановления пропущенных данных | 30 |
| 2.1.5 Пиктографики «Лица Чернова» | 30 |
| 2.1.6 Методы сокращения размерности многомерных данных | 33 |
| 2.2 Выбор инструмента решения..... | 35 |
| Глава 3 Решение поставленной задачи и описание полученных результатов | 41 |
| 3.1 Построение алгоритма информационной технологии | 41 |
| 3.2 Предварительная подготовка данных | 42 |
| 3.3 Визуальное исследование структуры пропущенных данных | 45 |
| 3.3.1 Идентификация недостающих данных..... | 46 |
| 3.3.2 Исследование закономерностей появления отсутствующих значений | 49 |
| 3.4 Метод множественного восстановления пропущенных данных | 51 |
| 3.5 Восстановление данных с использованием метода "bagImpute" | 54 |
| 3.6 Определение информативных признаков..... | 55 |
| 3.7 Построение визуальных образов | 56 |
| Глава 4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение..... | 59 |
| 4.1. Оценка коммерческого потенциала и перспективности проведения научных исследований | 59 |

| | |
|---|-----|
| 4.1.1. Потенциальные потребители результатов исследования | 59 |
| 4.1.2. Анализ конкурентных технических решений..... | 61 |
| 4.1.3. Технология QuaD..... | 62 |
| 4.1.4. SWOT-анализ | 63 |
| 4.2. Планирование научно-исследовательских работ | 65 |
| 4.2.1. Структура работ в рамках научного исследования..... | 65 |
| 4.2.2. Определение трудоемкости выполнения работ..... | 66 |
| 4.2.3. Разработка графика проведения научного исследования..... | 68 |
| 4.3. Бюджет научно-технического исследования (НТИ) | 74 |
| 4.3.1. Расчет материальных затрат НТИ..... | 74 |
| 4.3.2 Основная заработная плата исполнителей темы | 75 |
| 4.3.3 Дополнительная заработная плата | 77 |
| 4.3.4. Отчисления во внебюджетные фонды (страховые отчисления)..... | 77 |
| 4.3.5. Накладные расходы | 78 |
| 4.3.6. Формирование бюджета затрат научно-исследовательского проекта | 79 |
| 4.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования .. | 79 |
| Глава 5 Социальная ответственность..... | 84 |
| 5.1 Анализ выявленных вредных факторов | 84 |
| 5.2 Анализ психофизических факторов..... | 91 |
| 5.3 Анализ выявленных опасных факторов | 94 |
| 5.4 Экологическая безопасность..... | 97 |
| 5.5 Безопасность в чрезвычайных ситуациях | 97 |
| 5.6 Правовые и организационные вопросы обеспечения безопасности ... | 100 |
| Заключение | 102 |
| Список используемых источников..... | 104 |
| Приложение А | 110 |

Введение

Проблему избыточного веса называют одной из болезней цивилизации, поскольку ее распространенность в современном обществе постоянно растет. В последнее время статистические наблюдения фиксируют рост частоты ожирения у детей и подростков. По данным всемирной организации здравоохранения, за последние четыре десятилетия в мире в десять раз увеличилась доля детей и подростков (от 6 до 17 лет), страдающих ожирением [1].

В настоящее время существует множество подходов к лечению ожирения, накоплены массивы данных клинико-лабораторных показателей пациентов. В связи с этим, актуальной становится задача разработки информационных технологий для оперативной обработки этих массивов и получения новых знаний. Под информационной технологией понимается процесс, использующий совокупность средств и методов сбора, обработки и передачи данных для получения информации нового качества о состоянии объекта, процесса или явления.

Целью работы является создание информационной технологии оценки результатов лечения пациентов, страдающих ожирением разной степени, проводимого по пяти различным траекториям.

Выпускная квалификационная работа посвящена решению следующих задач:

- Анализ литературы и интернет-источников – для ознакомления с предметной областью и опытом решения подобных задач;
- Выбор методов и инструментов решения задачи;
- Оценка исходного массива клинико-лабораторных показателей – идентификация пропусков и исследование закономерностей появления отсутствующих значений;
- Подготовка данных к анализу – восстановление пропусков и формирование полного набора данных;

- Построение визуальных образов до и после лечения на основе восстановленных данных;
- Согласование полученных результатов со специалистами НИИ Курортологии и физиотерапии, формирование выводов;
- Расчет ресурсоэффективности и ресурсосбережения;
- Разработка комплекса мер технического, организационного, режимного и правового характера, минимизирующих негативные последствия разработки информационной технологии.

Результаты работы представлены в следующих научных публикациях и выступлениях на конференциях:

1. Публикация в журнале «Информационные технологии в науке, управлении, социальной сфере и медицине»: И.А. Лызин Использование визуализации при анализе медицинских данных // Информационные технологии в науке, управлении, социальной сфере и медицине. Сборник научных трудов IV международной научной конференции в 2 ч., Томск, 5-8 декабря 2017. - Томск: ТПУ С.339-342.

2. Участие в VIII международном молодёжном научном форуме «Новые форматы транснациональной научно-образовательной деятельности» с докладом «Информационная технология анализа результатов лечения детей с различными формами ожирения» 26 – 27 апрель 2018. ТГПУ диплом III степени.

3. IV Международная научно-практическая конференция молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук» (статья «Информационная технология анализа результатов лечения детей с различными формами ожирения» включена в работу конференции и принята в печать)

4. III Всероссийская конференция с международным участием «Здоровье и качество жизни» (статья «Разработка алгоритма анализа результатов лечения детей с эндокринопатиями» включена в работу конференции и принята в печать)

Обзор литературы

Проблеме обработки многомерных медицинских данных в настоящее время посвящено множество публикаций. Существует множество подходов к решению данной задачи – в зависимости от характера данных и поставленных целей. По данной теме как в России, так и за рубежом, публикуются статьи, в которых не только предлагаются новые методики анализа, но также подчёркивается важность темы. В таблице 1.1 представлен обзор статей основной темой которых является анализ многомерных медицинских данных с использованием различных методов.

Таблица 1.1 Анализ публикаций по теме исследования

| № | Авторы, название статьи, ссылка | Организация | Аннотация |
|---|---|--|--|
| 1 | Н. Самойленко, В. Кувина, С. Кувин Комплексный анализ медицинских данных [2]. | Воронежский государственный технический университет; Научный центр реконструктивно-восстановительной хирургии СО РАМН, г. Иркутск; Иркутская областная детская клиническая больница. | В статье излагаются методы комплексного анализа слабо формализованных данных в задачах выявления скрытых взаимодействий параметров (признаков) сложных систем, что актуально в задачах разработки средств программной поддержки принятия решений в системах автоматизированного проектирования сложных технических объектов, в автоматизированных системах профессиональной подготовки специалистов, а также в системах медицинской диагностики. |
| 2 | С.В. Кучерявский, И.А. Беляев Применение методов анализа многомерных данных для классификации клеток крови [3] | Алтайский государственный университет | В статье содержится описание методов анализа многомерных данных. Кроме этого статья содержит описание предложенного авторами алгоритма классификации способного давать приемлемый результат при работе с образцами различного качества и не зависящего от аппаратной части анализатора. |
| 3 | Д.Р. Струков, В.Л. Горохов Геоинформационные системы и многомерные статистические методы | ООО «Центр пространственных исследований» | В работе рассматривается модификация средств геостатистического анализа на основе ранговой «нормализации» данных. Описан алгоритм анализа |

| | | | |
|---|---|--|---|
| | пространственного анализа для исследования заболеваемости [4]. | | многомерных медицинских данных и показано их применение в практических задачах здравоохранения. |
| 4 | Осадчая И.А., Берестнева О.Г. Методы исследования структуры многомерных экспериментальных данных [5]. | Национальный исследовательский Томский политехнический университет | В работе рассматриваются особенности структурного анализа данных с использованием когнитивной графики и кластерного анализа. Описаны примеры решения двух прикладных задач анализа медицинских данных. |
| 5 | Лосев А.Г. Зенович А.В. Бочкарев О.А., Левшинский В.В. Интеллектуальный анализ многомерных термометрических данных в медицинской диагностике [6]. | Волгоградский государственный университет | В работе содержится описание технологии разработки консультативных интеллектуальных систем, а также методов искусственного интеллекта, предназначенных для анализа, моделирования и интерпретации медицинских термометрических данных. |
| 6 | Осадчая И. А., Берестнева О. Г., Немеров Е. В. Анализ многомерных медицинских данных с помощью пиктографиков «Лица Чернова» [7]. | Национальный исследовательский Томский политехнический университет | Основная идея работы заключается в том, что применение графики в исследовательских работах увеличивает скорость передачи информации и повышает уровень ее понимания. В статье дается описание метода визуализации медицинских данных с применением лиц Чернова. |
| 7 | Стромов Г.Г., Рыжков, Д.В. Исследование границ применимости методики интегральной оценки в задачах анализа многомерных медицинских данных [8]. | Национальный исследовательский Томский политехнический университет | Работа содержит описание общих подходов применения методики интегральной оценки состояния в задаче автоматизированного поиска областей интереса в трехмерных медицинских изображениях. |
| 8 | Романова, И.К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах [9] | МГТУ им. Н.Э. Баумана | В статье рассматриваются теоретические и практические аспекты решения проблемы визуализации многомерных данных как эффективного средства многомерного анализа систем. Описано несколько классификаций для методов визуализации в соответствии с объектами визуализации, типами данных, методом преобразования координат и данных. |

| | | | |
|----|--|--|---|
| 9 | Шаропин К. А. Берестнева О. Г. Воловоденко В. А Марухина О. В. Визуализация медицинских данных на базе пакета NOVOSPARK [10] | Национальный исследовательский Томский политехнический университет | Работа содержит описание различных подходов к визуализации результатов экспериментальных исследований. Представлены примеры решения прикладных задач с использованием NovoSpark Visualizer. |
| 10 | А.С. Герасимова Визуализация многомерных данных с помощью естественной поверхности [11]. | Алтайский государственный университет | В работе рассматривается новый подход визуализации многомерных данных использующий естественную поверхность (поверхность, содержащую все точки X_1, \dots, X_m вместе с криволинейной сеткой координат). |
| 11 | В.К. Финн, В.Г. Блинова, Е.С. Панкратова Е.Ф. Фабрикантова Интеллектуальные системы для анализа медицинских данных [12]. | Всероссийский институт научной и технической информации Российской Академии наук | Работа содержит описание применения метода качественного анализа данных посредством компьютерных интеллектуальных систем для задач медицинской диагностики. |
| 12 | Беребин М. А. Методология и практика разработки методик клинической (медицинской) психодиагностики на основе применения экспертного метода, методик обработки экспертных оценок и методов многомерного анализа данных [13]. | Южно-Уральский государственный университет | В работе представлено обоснование разработки методик на основе экспертного психодиагностического подхода, включающего клинический метод, методы сбора, обработки и многомерного математического анализа экспертных оценок. |
| 13 | А.Н. Кокоулин, Р.И. Южанинов Многомерный анализ данных по обращаемости в лечебные учреждения с помощью средств Oracle OLAP [14] | Пермский национально исследовательский политехнический университет | В статье описывается процесс анализа больших объемов медицинских данных с использованием технологии OLAP. Предложено дальнейшее развитие данной технологии, которое заключается в интеграции построенных и планируемых к реализации OLAP-кубов в Oracle Fusion Middleware. |
| 14 | М.Л. Никонорова Интеллектуальный анализ медицинских данных с использованием | Первый Санкт- Петербургский государственный медицинский университет им. академика | Основная идея данной работы посвящена проблеме эффективности применения информационных технологий в повседневной практической деятельности врача. Статья |

| | | | |
|----|---|---|---|
| | кейсовой технологии [15]. | И. П. Павлова | содержит описание модели принятия решения, позволяющая реализовать конкретную ситуацию. Кроме этого приведен алгоритм решения с использованием прикладного программного обеспечения RapidMiner. |
| 15 | А.И. Минжасова Статистический анализ медицинских данных [16]. | Омский государственный технический университет | В работе рассматриваются наиболее актуальные виды анализа: корреляционный, факторный, кластерный. Помимо этого, работа раскрывает сущность статистического анализа в области медицинских исследований. |
| 16 | В.В. Быкова, А.В. Катаева Методы и средства анализа информативности признаков при обработке медицинских данных [17]. | Сибирский федеральный университет | В статье подробно рассматриваются статистические методы оценки информативности признаков, используемые в медицинской диагностике: метод Шеннона, метод накопленных частот и метод Кульбака. Также в работе рассмотрены методы сокращения признакового пространства, такие как метод главных компонент и метод экстремальной группировки признаков |
| 17 | Кожевников С.Н., Денисов А.В., Новикова И.И., Ерофеев Ю.В. Использование многомерных методов статистического анализа для определения программ лечебно-профилактической помощи медицинским работникам в современных условиях [18] | Новосибирский НИИ гигиены | В работе рассмотрены и проанализированы данные аттестации рабочих мест, результаты социологического исследования. Результатом работы являются кластеры риска и патологии риска, определяющие закономерности и особенности формирования нарушений здоровья полученные в ходе применения методов многомерного анализа. |
| 18 | К. Ю. Куликова Статистический анализ медицинских данных при наличии пропусков [19]. | Санкт-Петербургский государственный университет | В работе рассматривается условная классификация типов пропусков и некоторые методы восстановления пропущенных данных. Приведен пример восстановления пропусков на медицинской базе, содержащей данные по 108 пациентам. |
| 19 | О. В. Марухина, О. Г. Берестнева, Е. Е. Мокина Визуализация и | Национальный исследовательский Томский | Работа содержит описание разработанного программного продукта на основе пиктографиков |

| | | | |
|----|--|--|---|
| | анализ многомерных экспериментальных данных [20] | политехнический университет | «Лица Чернова». Производиться оценка биологического системного состояния с помощью интегрированных индикаторов, а также методов обработки изображений пациентов с различными формами астмы. |
| 20 | Захаров А. А., Оленников Е. А., Паюсова Т. И., Петелина Т.И., Мусихина Н. А., Гапон Л. И., Осипова И. В., Такканд А. Г., Белослудцева О. Е. Научный анализ данных в медицинской информационной системе (на примере определения факторов, влияющих на уровень с-реактивного белка, с помощью нейронных сетей) [21]. | Тюменский государственный университет | В статье описываются подходы, которые позволяют использование медицинских информационных систем в качестве инструмента для клинических и биологических исследований. В качестве примера описывается технология, которая позволяет автоматизировать медицинские исследования, проводимые для определения факторов, влияющих на уровень маркера воспалительного ответа. |
| 21 | Осадчая И.А., Прокопьев Р.О. Технологии визуализации в задачах оценки функционального состояния [22]. | Национальный исследовательский Томский политехнический университет | Работа содержит описание применения различных методов визуализации экспериментальных медицинских данных (на примере исследования физиологических особенностей пациентов с различными формами бронхиальной астмы). |
| 22 | Jolita Bernatavičienė, Gintautas Dzemyda, Olga Kurasova, Virginijus Marcinkevičius, Viktor Medvedev The Problem of Visual Analysis of Multidimensional Medical Data [23]. | Biodynamics Research Unit, Mayo Foundation, Rochester | Рассматривается проблема визуального анализа многомерных медицинских данных. Также в статье описывается объединение таких методов анализа данных как классификации и визуализации. Этот синтез позволяет получить более объективные выводы по анализируемому данным. |
| 23 | R.A. Robb, D.P. Hanson, R.A. Karwoski, A.G. Larson, E.L. Workman, M.C. Stacy Analyze: A Comprehensive, operator-interactive software package for multidimensional medical image display and analysis [24]. | Institute of Mathematics and Informatics Vilnius Lithuania | Работа содержит описание разработанного комплексного программный пакет под названием ANALYZE, который позволяет детально исследовать и оценивать многомерные биомедицинские изображения. ANALYZE может использоваться с трехмерными методами визуализации на основе рентгеновской компьютерной томографии, радионуклидной эмиссионной томографии, |

| | | | |
|----|--|--|--|
| | | | ультразвуковой томографии и магнитно-резонансной томографии. Программное обеспечение полностью написано в «С» и работает на стандартных рабочих станциях UNIX. |
| 24 | Noor Elaiza, Abd Khalid, MarinaYusoff, Ezzatul Akma, Kamaru-Zaman Izyan Izzati, Kamsani Multidimensional Data Medical Dataset Using Interactive Visualization Star Coordinate Technique [25]. | University Technology MARA Shah Alam, Malaysia | В этой статье исследуется возможность интерактивного метода визуализации координаты для определения корреляции кластеров между выбранными атрибутами с использованием интерактивной координаты звезд для многомерных наборов данных. |
| 25 | Anbarasi M.S. Outlier Detection for Multidimensional Medical Data [26]. | Pondicherry Engineering College | Основная тема статьи посвящена проблеме обнаружения выбросов в медицинских данных. Это важная проблема исследования, целью которой является найти объекты, которые значительно отличаются друг от друга |

На основе проведенного обзора литературы сделан вывод об актуальности разрабатываемой ИТ анализа результатов лечения детей с различными формами ожирения. Прикладная значимость заключается в возможности комплексного анализа результатов лечения пациентов и рекомендации по выбору метода лечения.

Глава 1 Постановка задачи

1.1 Информационные технологии анализа данных

Анализ данных – это процесс проверки, очистки, преобразования и моделирования данных с целью выявления полезной информации, предложения и поддержки принятия решений. Анализ данных имеет множество аспектов и подходов, охватывающих различные методы под разными названиями, в разных областях бизнеса и социальных наук.

Под многомерными данными понимаются данные содержащие информацию о двух или более признаках для каждого изучаемого объекта.

Многомерные данные характеризуются:

- Большой размерность признакового пространства, вызванной наличием нескольких десятков характеристик;
- Отсутствие представительной статистики, которую зачастую невозможно получить из-за больших материальных затрат на их получение;
- Наличием пропусков в данных;
- Интервальным характером данных, обусловленным неопределенностью условий их получения, статистические методы принципиально не могут оперировать с указанными данными.

Анализ многомерных данных (АМД) – это подход, основанный на применении математических методов, в котором исследуются явления, процессы и системы, характеризующие совокупностью величин (переменных).

Словосочетание «медицинские данные» используется как в узком, так и в широком смысле. Медицинские данные в широком смысле – это любые данные, относящиеся к медицине в целом. В узком смысле – это данные, относящиеся непосредственно к человеку как к пациенту, то есть информация о его здоровье, особенностях организма. В последнее время термин всё чаще употребляется именно в узком смысле. Далее будем придерживаться именно такого понимания медицинской информации.

Работа с медицинскими данными имеет свои особенности. При работе с медицинскими данными необходимо: не допускать потерь информации – использовать данные с той точностью, которая имеет место при измерении; с другой – при представлении результатов статистической обработки данных не приводить избыточной информации – в большинстве случаев достаточно той точности представления информации, что и в исходных данных. Обычно при предъявлении числовых данных указываются два знака после запятой.

При наличии пропущенных данных нельзя использовать так называемое «обнуление», т.е. приписывать кодовое число нуль, так как это в большинстве случаев совпадает с кодированием нормы по данному признаку. Также методически неверно использовать среднее по классу, особенно при малых выборках, так как классы далеко не всегда являются однородными.

Сложность многомерных медицинских данных обусловлена следующими особенностями:

- Плохая структурированность;
- Избыточность информации о пациенте;
- Недостаток достоверной информации;
- Трудность формализации данных.

Для анализа данных в том числе и медицинских используются информационные технологии. Информационные технологии включают два фактора – машинный и человеческий. Конкретным воплощением информационных технологий в основном выступают автоматизированные системы, и лишь в этом случае принято говорить о компьютерных технологиях.

Авторами статьи «Классификация информационных технологий» представлена полная классификация информационных технологий [27] с разделением на 7 уровней. На рис. 1.1 представлена полная классификация информационных технологий. Создаваемая информационная технология анализа медицинских данных соответствует пятому уровню «технологии обработки данных и извлечения знаний для принятия решений», подуровню

технология извлечения знаний из данных «Data Science (KDD, ETL, Data Mining)».



Рисунок 1.1 Классификация информационных технологий

1.2 Описание предметной области

Ожирение у детей и подростков является актуальной и весьма серьезной проблемой. Распространенность, которой возрастает: в развитых странах мира 25% подростков имеют избыточную массу тела, а 15% страдают ожирением.

Избыточный вес в детстве – значимый предиктор ожирения во взрослом возрасте: 50% детей, имевших избыточный вес в 6 лет, становятся тучными во взрослом возрасте, а в подростковом возрасте эта вероятность увеличивается до 80%. Избыточный вес и ожирение – это результат формирования аномальных или чрезмерных жировых отложений, которые могут наносить вред здоровью.

Существует два основных фактора, которые могут привести к детскому ожирению: алиментарный и эндокринный.

Алиментарный фактор связан с неправильным питанием или малоподвижным образом жизни. Ребенок переедает, но не расходует накопленную энергию. Эндокринный фактор связан в первую очередь с нарушением работы эндокринной системы, которая отвечает за работу внутренних органов. К ожирению могут привести следующие сбои:

- Генетическая предрасположенность. Ожирение у обоих родителей дает 80%-ю гарантию того, что такая же болезнь разовьется и у ребенка;
- Гормональные нарушения. Это нарушение работы желез внутренней секреции: болезни надпочечников, поджелудочной железы, пониженная активность щитовидки, детский гипотиреоз – заболевания, которые нередко приводят к ожирению.

Значимость проблемы ожирения определяется угрозой инвалидизации пациентов молодого возраста и снижением общей продолжительности жизни в связи с частым развитием тяжелых сопутствующих заболеваний. К ним можно отнести: сахарный диабет 2 типа, артериальную гипертонию, дислипидемию, атеросклероз и связанные с ним заболевания, желчекаменную болезнь, остеохондроз. Ожирение снижает устойчивость к инфекционным и простудным заболеваниям, кроме того, резко увеличивает риск осложнений при оперативных вмешательствах и травме [28].

В рамках сотрудничества с НИИ курортологии и физиотерапии г. Томска нами был получен массив данных, содержащий значения групп клинико-лабораторных показателей пациентов. Состояние каждого пациента

оценивалось при поступлении на лечение и по окончании курса реабилитации. Пациенты – дети и подростки в возрасте от 6 до 17 лет, с проблемами веса. База состоит из 276 пациентов, поделенных на 5 групп и 106 показателей. Группы характеризуют пациентов, проходящих лечение по разным методикам.

Имеются следующие основные клинико-лабораторные показатели:

ОТ – объем талии пациента

ОБ – объем бедер

Масса, кг – масса тела в кг

Избыток, % – избыток массы тела в %

ИМТ – индекс массы тела

САД, мм.рт.ст. – систолическое артериальное давление

ДАД мм рт. ст. – диастолическое артериальное давление

НОМА (у.е.) – индекс инсулинорезистентности Хома

ТФН – толерантность к физической нагрузке

Дв. Пр. – двойное произведение (насыщение миокарда кислородом)

Общая р-ть – общая работоспособность

Общие липиды (4-8 г/л) – общие липиды крови

ТАГ (менее 1,7ммоль/л) – содержание триацилглицеридов в сыворотке крови

Холестерин (ОХС) (менее 4 ммоль/л) – содержание общего т холестерина в сыворотке крови

аХС (ЛПВП) (более 1,03 ммоль/л) – содержание альфа-холестерина в сыворотке крови

ЛПОНП (ммоль/л) – липопротеиды очень низкой плотности

ЛПНП (ммоль/л) – липопротеиды низкой плотности

Индекс атерогенности – индекс атерогенности

Щелочная фосфатаза (ед.) – щелочная фосфатаза в сыворотке крови

Каталаза (мкатал/л) – каталаза сыворотки крови

МДА – Малоновый диальдегид в сыворотке крови

Сиаловые кислоты – содержание сиаловых кислот в сыворотке крови

Церулоплазмин – содержание церулоплазмينا в сыворотке

Глюкоза (менее 5,6ммоль/л) уровень глюкозы крови натощак

T3 (1,0-2,8) – содержание трийодтиронина в сыворотке крови

T3св (4-8.6) – содержание свободного трийодтиронина

T4 (53-158) – содержание тетраiodтиронина

T4св (10-23.2) – содержание свободного тетраiodтиронина

ТТГ (0,23-3,4) – содержание тиреотропного гормона в сыворотке крови

АТ ТПО (до 30 у.е.) – антитела к тиреопероксидазе в сыворотке крови

Инсулин – содержание инсулина в сыворотке крови

Кортизол – содержание кортизола в сыворотке крови

Лептин – содержание лептина в сыворотке крови

Серотонин – содержание серотонина в сыворотке крови

ИЛ-6 – содержание интерлейкина 6 в сыворотке крови

ФНО (не >2,5 пг/мл) – содержание фактора некроза опухоли в сыворотке крови

IgA (0,7-3,0 г/л) – концентрация иммуноглобулина А в сыворотке крови

IgG (8,0-16,0 г/л) – концентрация иммуноглобулина G

IgM (0,8-2,5 г/л) – концентрация иммуноглобулина M

ЦИК (40-100 у.е.) – циркулирующие иммунные комплексы в сыворотке крови

Лизоцим (40-60 у.е.) – активность лизоцима в сыворотке крови

T-лимфоциты (45-70) % – циркулирующие иммунные комплексы в сыворотке крови

T-хелперы (33-50%) – циркулирующие иммунные комплексы в сыворотке крови

T-супрессоры (16-39 %) – циркулирующие иммунные комплексы в сыворотке крови

B-лимфоциты (6-23%) – циркулирующие иммунные комплексы в сыворотке крови

КК – уровень каллекриина в сыворотке крови

ПИ – протромбиновый индекс в сыворотке крови

МГ – макроглобулин в сыворотке крови

АПФ – ангиотензинпревращающий фермент в сыворотке крови

NO – содержание оксида азота в крови

ОК ЛПНП – окисление липопротеидов низкой плотности в сыворотке
крови

Задача исследования заключается в построении визуальных образов групп пациентов по и после лечения для оценки его эффективности.

Глава 2 Выбор методов решения задачи

2.1 Выбор технологии и описание методов решения задачи

2.1.1 Технология Data Mining

Термин «Data Mining» на сегодняшний день не имеет устоявшегося перевода. Для передачи смысла обычно используют следующие словосочетания: добыча данных: извлечение данных, просев информации, а также интеллектуальный анализ данных. Более полным и точным переводом является словосочетание «обнаружение знаний в базах данных». Термин введён Григорием Пятецким-Шапиро в 1989 году.

Data Mining – это процесс основная цель которого заключается в извлечении информации из набора данных и преобразовании ее в понятную структуру для дальнейшего использования.

Основу методов Data Mining составляют различные методы моделирования, классификации и прогнозирования, основанные на применении искусственных нейронных сетей, деревьев решений, эволюционного программирования, ассоциативной памяти, нечёткой логики, в том числе алгоритмы k -средних и k -медианы; методы поиска ассоциативных правил, метод ограниченного перебора, генетические алгоритмы и разнообразные методы визуализации данных. К методам Data Mining нередко относят статистические методы такие как: дескриптивный анализ, корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов и множество других методов

После изучения методов и алгоритмов, для решения поставленной задачи восстановления пропущенных данных будет использован метод множественного восстановления пропущенных данных, но также будет рассмотрен метод восстановления данных на основе деревьев регрессий. В качестве методов сокращения размерности будут рассмотрены метод главных компонент, факторный анализ и метод исключения неинформативных

признаков. В качестве средства визуализации данных были выбраны лица Чернова.

2.1.2 Анализ выбросов

Очистка данных – это процесс исправления выбросов и другой неправильной и нежелательной информации. Существует несколько типов процесса очистки данных, которые зависят от типа данных, подлежащих очистке. Для методов количественных данных обнаружение выбросов может быть использовано для устранения аномалии в данных.

Под "выбросом" понимается наблюдение, которое "слишком" велико или "слишком" мало по сравнению с большинством других имеющихся наблюдений. Обычно для обнаружения выбросов используют диаграмму размахов. Диаграмма размаха, также известная как ящик с усами. Этот график упрощает данные, оставляя только несколько параметров: медиану, верхнюю и нижние квартили (т.е. такие числа, что четверть данных больше верхней квартили и четверть меньше нижней), существенный минимум и максимум (наименьшее и наибольшее значения, не считая выбросов), а также выбросы, которые существенно меньше или больше остальных чисел в ряду. Обычно этот способ применяется, если необходимо визуально сравнить несколько категорий данных.

2.1.3 Анализ пропусков

Когда в исходном наборе присутствуют пропуски необходимо оценить причину их появления и их влияние на дальнейший анализ. Если пропущенные значения не обрабатываются должным образом в конечном итоге могут получиться результаты не соответствующие действительности. Поэтому анализ пропущенных значений является важной задачей.

Существует три типа недостающих данных:

- Полностью случайные пропуски (*missing completely at random MCAR*).

Данные являются MCAR, если вероятность пропуска данных в переменной X

не связана с другими переменными и значениями самих X . Другими словами отсутствие данных полностью несистематично;

- Случайные пропуски (*missing at random MAR*). Данные типа MAR встречаются, когда пропуски во всей совокупности данных случайно распределены не по всем переменным, а внутри определенных подгрупп этих переменных;

- Отсутствуют не случайно (*non-ignorable missingness MNAR*). Данные являются MNAR, если вероятность пропуска данных систематично связана с предполагаемыми значениями этих пропущенных данных. Таким образом, такие данные отсутствуют из-за их значений. Например, при медицинском осмотре взвешивали только людей с избыточным весом.

Слишком большое количество пропущенных значений является проблемой для дальнейшей обработки и анализа данных. Обычно максимальный порог пропущенных данных составляет 15% от общего количества выборки. Если пропущенные данные превышают 15%, необходимо отказаться от этого показателя.

Методы обработки пропущенных значений:

Complete case analysis этот метод предполагает исключение из анализа показатели в которых имеются пропуски. Данный метод наиболее часто используемый в программах статистического анализа, таких как SPSS и SAS для обработки отсутствующих значений.

Available case analysis – метод основанный на игнорировании недостающих значений. Это позволяет увеличить размер выборки, однако этот подход приводит к тому, что размеры выборки изменяются между переменными, используемыми в анализе.

Imputation analysis – предполагает восстановление пропущенных данных. Восстановление данных осуществляется с использованием явного или неявного подхода к моделированию. Явный подход к моделированию предполагает, что переменные имеют определенное предсказательное распределение и оценивают параметры каждого распределения, которое

используется для восстановления. Данный подход заключается в восстановлении данных при помощи среднего, медианы, регрессии и предположению о распределении. Неявный подход к моделированию сфокусирован на вычислении алгоритма, необходимого для получения точных значений восстановления.

2.1.4 Метод множественного восстановления пропущенных данных

Метод множественного восстановления пропущенных данных (*Multiple Imputation, MI*) – это способ заполнения пропусков при помощи повторного моделирования. Множественное восстановление часто применяется для работы с пропущенными данными в сложных ситуациях. При этом подходе из существующего набора данных с пропущенными значениями создается несколько полных наборов данных (обычно от трех до десяти). Для замещения пропущенных значений в производных наборах данных используются методы Монте-Карло.

К каждому из производных наборов данных применяются стандартные статистические методы, а на основании их результатов формируются оценки окончательных результатов и доверительные интервалы, которые учитывают неопределенность, созданную пропущенными значениями.

2.1.5 Пиктографики «Лица Чернова»

Визуализация является неотъемлемой частью анализа данных. Визуализация данных – это представление числовой и текстовой информации в виде графиков, диаграмм, карт, структурных схем, посредством которых обеспечивается наиболее эффективная работа человека по их изучению [29]. Визуализация данных, как правило, используются для широкой популяризации данных и облегчения их восприятия неспециалистами. Помимо всего перечисленного визуализация используются для обобщения и анализа статистических данных.

Использование визуализация при анализе данных оказывает следующие преимущества:

- Акцентирование внимания на разных аспектах данных;
- Анализ большого набора данных со сложной структурой;
- Удерживание внимания человека и уменьшение его информационной перегрузки;
- Однозначность и ясность выводимых данных;
- Выявление взаимосвязей и отношений, содержащихся в информации.

К способам визуального или графического представления данных относят структурные схемы, графики, диаграммы, карты, таблицы, отчеты, списки и т.д. Методы визуального представления данных могут находить следующее применение:

- Представлять пользователю информацию в наглядном виде;
- Компактно описывать закономерности, присущие исходному набору данных;
- Сжимать информацию или снижать размерность;
- Находить выбросы и шумы в исходных наборах данных.

С помощью графиков легче уяснить закономерности развития, распределения и размещения явлений и можно сделать выводы, которые на базе табличного материала были бы затруднительными. Восприятие визуальной информации физиологически, является основной для человека. Существуют многочисленные научные исследования, доказывающие, что:

- Человек воспринимает через зрение 90% информации;
- В глазах находятся 70% сенсорных рецепторов;
- Около половины нейронов головного мозга человека задействовано в обработке визуальной информации;
- У человека, работающего с визуальной информацией на 17% выше производительность.

Подробное описание данных исследований приводиться в книге «Визуализация данных. Бизнес-аналитика. От данных к знаниям» (Паклин Н. Б., Орешков В. И.) [30].

Многомерные пиктографики являются не очень простым, но мощным исследовательским инструментом разведочного анализа данных. Основная идея такого метода анализа базируется на человеческой способности "автоматически" фиксировать сложные связи между многими переменными, если они проявляются в последовательности элементов. Основная идея использования пиктографиков заключается в представлении отдельных наблюдений в виде некоторых графических объектов, где значения переменных соответствуют размерам этих объектов или определенным свойствам (одно наблюдение = один объект). Это соответствие таково, что внешний вид объекта изменяется в зависимости от набора значений. Анализ информации при помощи такого способа отображения основан на способности человека интуитивно находить сходства и различия в чертах объекта. Конкретную природу проявившихся взаимосвязей между переменными позволяет выявить уже последующий анализ данных, основанный на изучении этого интуитивно обнаруженного сходства.

В качестве средства построения визуальных образов был выбран такой тип пиктографиков как лица Чернова.

Лица Чернова (*Chernoff Faces*), разработанные Herman Chernoff в 1973 году, являются наиболее интересным типом пиктографиков. Лица Чернова – это схема визуального представления мультивариативных данных в виде человеческого лица. Каждая часть лица: глаза, нос, рот – представляет собой значение определенной переменной, назначенной для этой части (всего 18) [31].

Основная идея использования лиц заключается в том, что люди легко распознают лица и без труда замечают небольшие изменения. Для человека очень естественно смотреть на лица, так как все люди делают это каждый день. Легко делать сравнения и легко выявлять отклонения.

2.1.6 Методы сокращения размерности многомерных данных

Сокращение размерности может проводиться несколькими методами: методом главных компонент, факторным анализом, отбрасыванием неинформативных признаков. Рассмотрим более подробно каждый из этих методов.

Анализ главных компонент (*principal components analysis, PCA*) – это способ снижения размерности данных, который преобразует большое число скоррелированных переменных в гораздо меньший набор нескоррелированных переменных, называемых главными компонентами.

Первая главная компонента определяется по формуле 2.1:

$$PC_1 = a_1X_1 + a_2X_2 + \dots + a_kX_k, \quad (2.1)$$

это взвешенная комбинация k наблюдаемых переменных, которая учитывает наибольшую дисперсию (долю изменчивости) исходного набора переменных. Вторая главная компонента – это линейная комбинация, которая учитывает наибольшую дисперсию исходного набора переменных при условии, что она ортогональна первой главной компоненте (то есть не коррелирует с ней). Каждая следующая компонента учитывает наибольшую возможную дисперсию, оставаясь нескоррелированной с остальными компонентами. Теоретически можно выделить столько главных компонент, сколько переменных. Однако из практических соображений предполагается, что можно охарактеризовать полный набор переменных посредством гораздо меньшего набора компонент.

При использовании данного метода возникает проблема интерпретации новых переменных, следовательно, данный метод не подходит для решения поставленной задачи.

Факторный анализ (*factor analysis, FA*) – это статистический метод, используемый для описания изменчивости среди наблюдаемых коррелированных переменных в терминах потенциально меньшего числа ненаблюдаемых переменных.

Цель FA – объяснить корреляции внутри набора наблюдаемых переменных меньшим набором более фундаментальных ненаблюдаемых переменных, которые лежат в основе данных. Эти гипотетические ненаблюдаемые переменные называют факторами. Каждый фактор должен объяснять дисперсию, разделенную между двумя или более наблюдаемыми переменными, так что формально они называются общими факторами. Модель выглядит так 2.2:

$$X_i = a_1 F_1 + a_2 F_2 + \dots + a_p F_p + U_i, \quad (2.2)$$

где X_i – это i -ая наблюдаемая переменная ($i = 1 \dots k$),

F_j – это общие факторы ($j = 1 \dots p$), и $p < k$.

U_i – это уникальная составляющая переменной

X_i (не объясненная общими факторами). Параметр a_1 можно интерпретировать как степень вклада каждого фактора в наблюдаемую переменную [32].

Метод Кульбака

Существуют большое количество методов определения информативности признаков. Однако в отличие от других критериев статистической значимости различий, мера Кульбака позволяет оценить не достоверность различий между распределениями, а степень этих различий. Метод анализа признаков путем оценки информативности с использованием информативной меры Кульбака получил широкое применение в медицине при рассмотрении отдельных факторов, влияющих на постановку диагноза.

Метод Кульбака – предлагает в качестве оценки информативности меру расхождения между двумя классами, которая называется дивергенцией. Согласно этому методу информативность или дивергенция Кульбака вычисляется по формуле 2.3 [33]:

$$I(x_j) = \sum_{i=1}^G [P_{i1} - P_{i2}] \cdot \log_2 \frac{P_{i1}}{P_{i2}}, \quad (2.3)$$

где G – число градаций признака;

P_{i1} – появление i -ой градации в первом классе вычисляется по формуле 2.4.

$$P_{i1} = \frac{m_{i1}}{\sum_{i=1}^G m_{i1}}, \quad (2.4)$$

где m_{i1} – частота появления i -ой градации в первом классе; Знаменатель – появление всех градаций в первом классе, то есть общее число наблюдений в первом классе.

P_{i2} – появление i -ой градации во втором классе вычисляется по формуле 2.5.

$$P_{i2} = \frac{m_{i2}}{\sum_{i=1}^G m_{i2}}, \quad (2.5)$$

где m_{i2} – частота появления i -ой градации во втором классе; Знаменатель – появление всех градаций во втором классе, то есть общее число наблюдений во втором классе.

Для решения задачи сокращения размерности был выбран метод Кульбака.

2.2 Выбор инструмента решения

На данный момент насчитываются десятки качественных статистических пакетов с возможностью визуализации данных. Рассмотрим наиболее популярные из них.

1. R – язык программирования для статистической обработки данных и работы с графикой, а также бесплатное программное обеспечение с открытым исходным кодом (чтение: создание визуализации данных бесплатно). Язык разрабатывался как аналогичный языку S и является его альтернативной реализацией, хотя между языками есть существенные отличия, но в большинстве своём код на языке S работает в среде R. Изначально R был разработан сотрудниками статистического факультета Оклендского университета Россом Айхэкой и Робертом Джентлменом. Язык

и среда поддерживаются и развиваются организацией R Foundation. R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ. Распространяется в виде откомпилированных приложений под ряд операционных систем: FreeBSD, Solaris и другие дистрибутивы Unix и Linux, Mac OS X, Microsoft Windows. Кроме того, можно просмотреть исходный код, изменить его и перераспределить его по общедоступной лицензии GNU.

Преимущества визуализации данных с использованием R:

- Язык R был разработан специально для анализа данных. Визуализация данных является ключевым компонентом анализа данных. Таким образом, язык позволяет легко преобразовывать необработанные данные в профессиональные визуализации данных с использованием лучших отраслевых практик.

- R является модульным и расширяемым. Это означает, что можно легко расширить возможности визуализации данных R всего несколькими строками кода. R может автоматически загружать, устанавливать пакеты визуализации данных в память во время выполнения. Пакеты – это библиотеки для работы специфических функций. Это означает, что есть огромное количество возможностей при создании визуализации данных с R.

- R предоставляет несколько способов публикации и развертывания визуализации данных. Например, можно программно экспортировать визуализацию данных в виде файлов JPG, PNG, PDF или SVG. Также можно использовать серверные сценарии или веб-службы для визуализации данных для отчетов и приложений. Кроме того, можно создавать интерактивные визуализации интерактивных данных с использованием таких фреймворков, как Shiny.

Для удобства работы с R существует ряд графических интерфейсов, таких как: RStudio, SciViews-R, Statistical Lab, JGR, RKWard, R Commander, Rattle. Кроме того, в ряде текстовых и кодовых редакторов предусмотрены специальные режимы для работы с R, в частности в ConTEXT, Emacs (Emacs

Speaks Statistics), jEdit, Kate, Syn, TextMate, Tinn-R, Vim, Bluefish, WinEdt (с пакетом RWinEdt), Gedit (с пакетом rgedit/gedit-r-plugin) [34].

2. Язык Python был создан Гвидо ван Россумом в 1991 году. Python – это высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода. Синтаксис ядра Python минималистичен. В то же время стандартная библиотека включает большой объём полезных функций.

В этом языке делается акцент на производительности и удобочитаемости кода. Python поддерживает несколько парадигм программирования, в том числе объектно-ориентированное, структурное, функциональное, аспектно-ориентированное и императивное. Код в Python организовывается в функции и классы, которые могут объединяться в модули.

Python распространяется под свободной лицензией Python Software Foundation License, позволяющей использовать его без ограничений в любых приложениях, включая проприетарные [35]. Есть реализации интерпретаторов для JVM (с возможностью компиляции), MSIL (с возможностью компиляции), LLVM и других. Проект PyPy предлагает реализацию Python с использованием JIT-компиляции, которая значительно увеличивает скорость выполнения Python-программ.

Python – активно развивающийся язык программирования, новые версии выходят примерно раз в два с половиной года. Вследствие этого и некоторых других причин на Python отсутствуют стандарт ANSI, ISO или другие официальные стандарты, их роль выполняет CPython.

3. MATLAB – это высокоуровневый язык и интерактивная среда для программирования, визуализации результатов и численных расчетов. MATLAB был создан Клив Моллером, профессором компьютерных наук Университета Нью-Мексико, в 1970-х годах. С помощью MATLAB можно создавать модели и приложения, разрабатывать алгоритмы, анализировать данные. Использование MATLAB включает в себя матричные вычисления, разработку и выполнение алгоритмов, создание пользовательских

интерфейсов и визуализацию данных. MATLAB используется инженерами и учеными во многих областях, таких как обработка изображений и сигналов, связь, системы управления для промышленности, интеллектуальная конструкция сетки, робототехника, а также компьютерное финансирование. [36]

4. Программа SPSS (*Statistical Package for the Social Sciences*), относительно недавно стала принадлежать компании IBM и сменила название на PASW (*Predictive Analytics SoftWare*). Система SPSS обладает широкими графическими возможностями. Она включает в себя большое количество разнообразных категорий и типов графиков, в том числе научные, деловые, трехмерные и двухмерные графики в различных системах координат, специализированные статистические графики – гистограммы, матричные, категоризованные графики и др. Для работы SPSS необходимо не менее 1 Гб оперативной памяти [37].

5. Развитие SAS началось в 1966 году в Университете Северной Каролины. Первоначально система была предназначена для анализа сельскохозяйственных данных. Сегодня SAS продолжает развиваться в компании SAS Institute. Область применения SAS – самые разнообразные научные исследования, бизнес аналитика и т. д. SAS – это программный пакет, который может обрабатывать, изменять, управлять и извлекать данные из различных источников и выполнять статистический анализ. В системе SAS реализован собственный язык программирования. Взаимодействие с программой возможно, как в консольном режиме, так и через графический интерфейс, который представляет собой графическую оболочку для упрощенного ввода команд языка программирования SAS [38].

6. RapidMiner – это программная платформа основная идея которой заключается в том, что аналитик не должен программировать при выполнении своей работы.

RapidMiner, ранее известная как YALE, была разработана в 2001 года Ральфом Клинкаенбергом, Ингом Мерсвой и Саймоном Фишером в

подразделении искусственного интеллекта Технического университета Дортмунда. В 2007 году название программного обеспечения было изменено с YALE на RapidMiner.

RapidMiner обеспечивает 99% расширенного аналитического решения на основе шаблонов, которые ускоряют доставку и уменьшают количество ошибок, почти устраняя необходимость писать код. RapidMiner предоставляет методы интеллектуального анализа данных и машинного обучения, включая: загрузку и преобразование данных, предварительная обработка данных и визуализация, интеллектуальная аналитика и статистическое моделирование, оценка и развертывание.

RapidMiner написан на языке программирования Java. RapidMiner предоставляет графический интерфейс для проектирования и выполнения аналитических рабочих процессов. Рабочие процессы в RapidMiner состоят из нескольких «Операторов». Каждый оператор выполняет одну задачу внутри процесса, а выход каждого оператора формирует вход следующего. Отдельные функции можно вызывать из командной строки. RapidMiner предоставляет схемы обучения, модели и алгоритмы и может быть расширен с использованием сценариев R и Python.

Функцию RapidMiner можно расширить с помощью дополнительных плагинов, которые доступны через RapidMiner Marketplace [39].

Таблица 2.1 – Сравнительный анализ инструментов решений

| Признак/Инструмент | R | Python | MATLAB | SAS | Stata | RapidMiner |
|---------------------------------|---|--------|--------|-----|-------|------------|
| Наличие графического интерфейса | + | + | + | + | + | + |
| Наличие консольного ввода | + | + | + | + | + | - |
| Свободное распространение | + | + | - | - | - | -/+ |
| Работа с графикой | + | + | + | + | + | + |
| Поддержка ОС Linux и Windows | + | + | - | + | + | + |
| Простота в понимании и изучении | + | + | - | + | + | + |

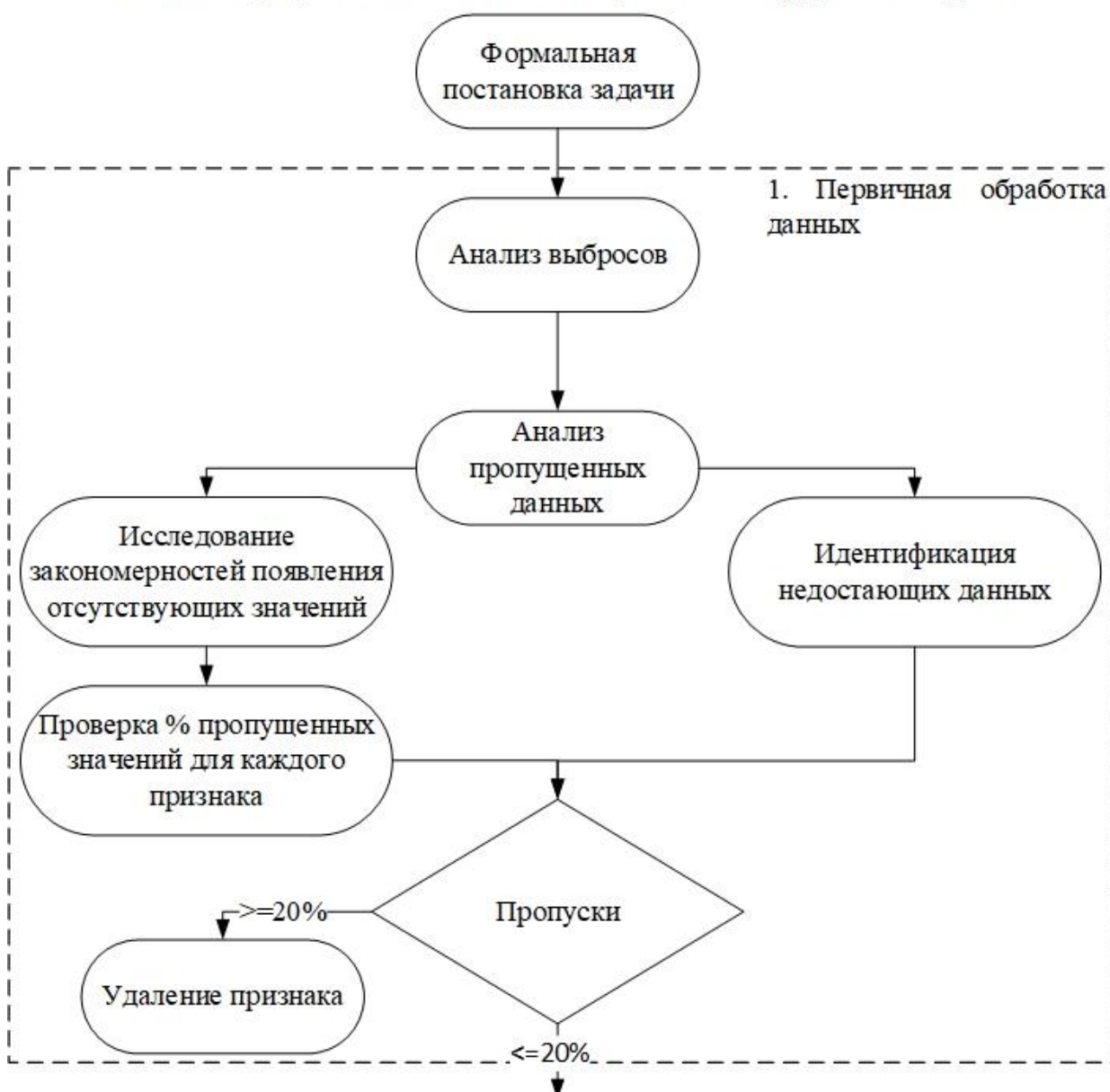
В качестве инструмента решения поставленной задачи был выбран язык R, так как R является популярным инструментом для визуализации данных. Язык R является одним из ведущих статистических инструментов в мире. Он активно применяется в генетике, молекулярной биологии и биоинформатике, науках об окружающей среде (экология, метеорология) и сельскохозяйственных дисциплинах. Также R все больше используется в обработке медицинских данных. В качестве среды разработки выбрано – Rstudio. RStudio – свободная среда разработки программного обеспечения с открытым исходным кодом для языка программирования R.

Глава 3 Решение поставленной задачи и описание полученных результатов

3.1 Построение алгоритма информационной технологии

Был разработан алгоритм информационной технологии, позволяющий осуществить восстановление пропусков в данных и комплексно оценить результаты лечения пациентов (рисунок 3.1). Алгоритм предполагает в качестве обработки клинико-лабораторных показателей удаление выбросов, выполнено восстановление пропусков в данных и построены визуальные образы с применением пакетов языка R.

ИТ анализа результатов лечения детей с различными формами ожирения



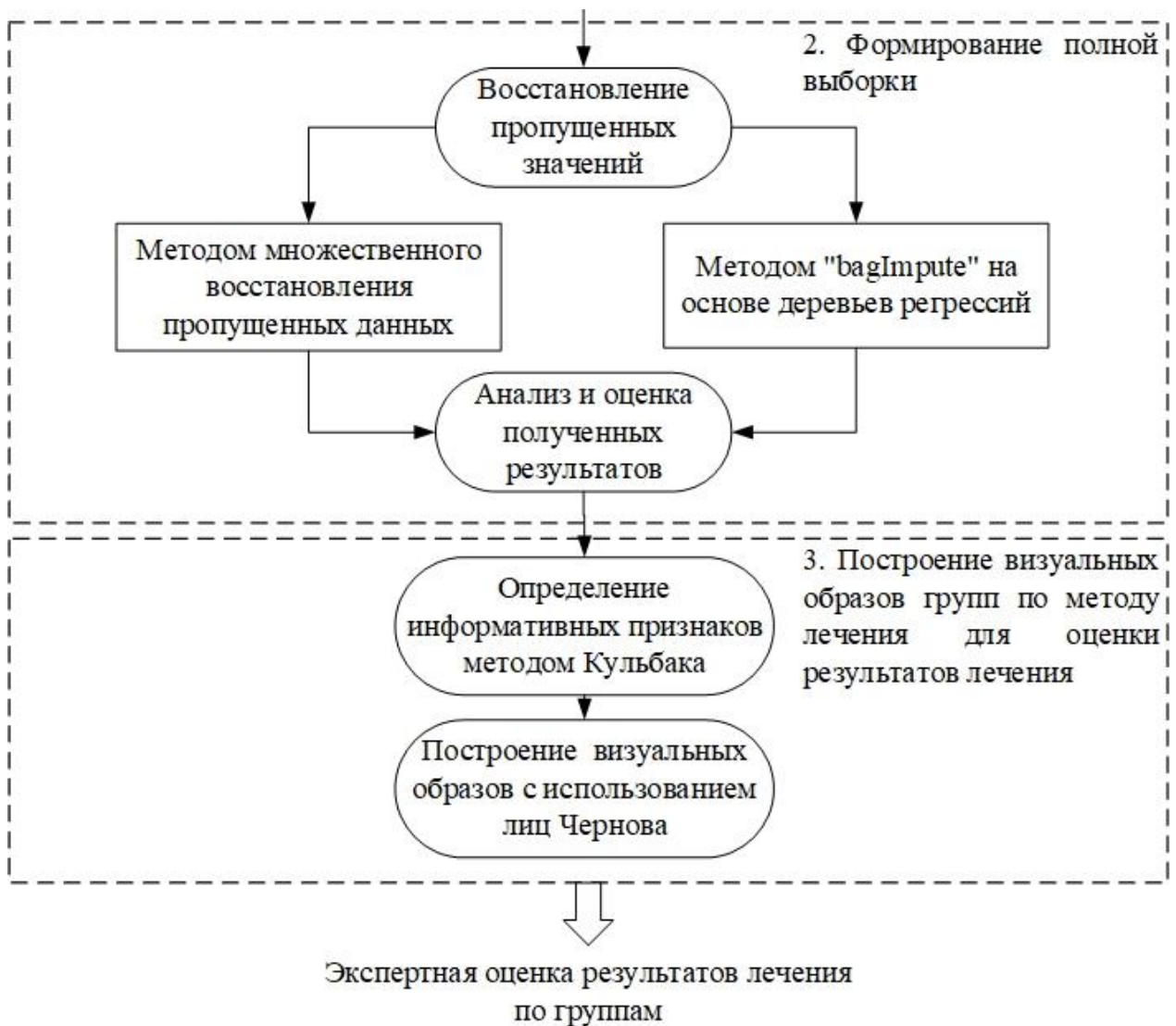


Рисунок 3.1 Алгоритм информационной технологии

3.2 Предварительная подготовка данных

Анализ целесообразно начать с подготовки исходных данных и проверки исходного массива на выбросы. Исходные данные располагаются в таблице Excel.

Анализ выбросов – это важный этап в подготовке исходных данных к восстановлению позволяющий избежать отклонения восстановленных данных от разумных значений. Удаление выбросов осуществлялось на основании таблицы 3.1 в которой содержится список лабораторных показателей и границы нормы для каждого из них.

Для анализа выбросов была использована диаграмма размахов. В R этот график реализован с помощью обычной функции `boxplot`. В работе используется 106 показателей, поэтому для большей наглядности представлен график `boxplot` для 10 показателей.

Функция `boxplot()` не только рисует картинку, но и сохраняет все ее параметры в объекте. Из которого можно извлечь их.

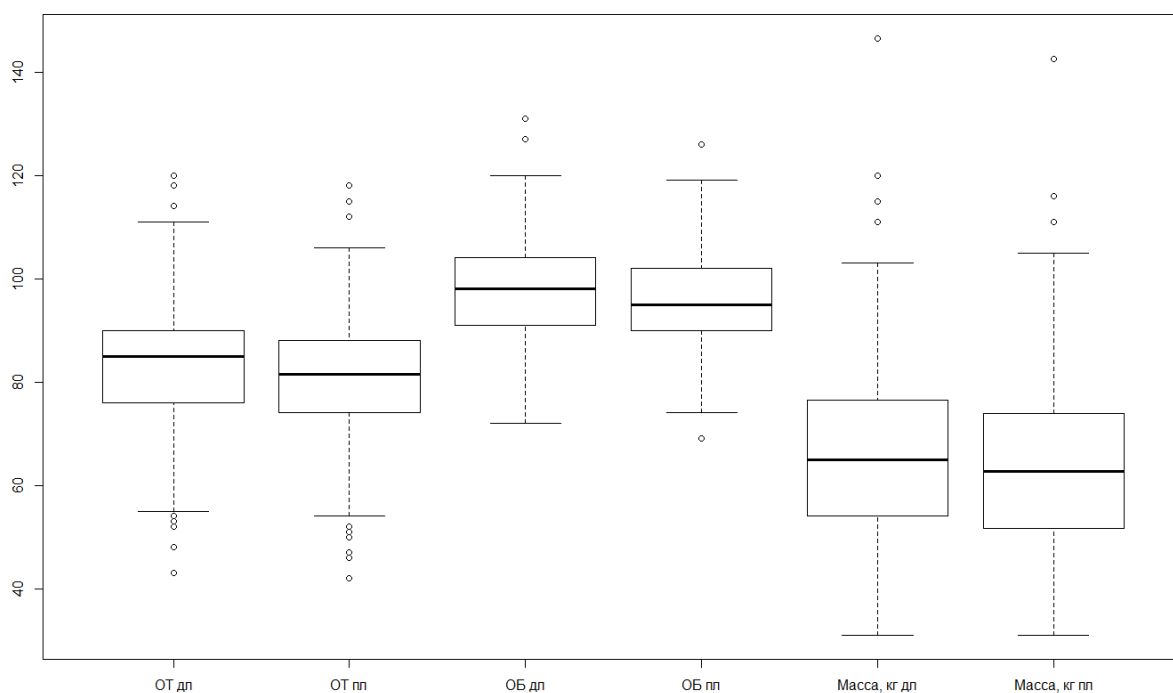


Рисунок 3.2 Диаграмма размаха для 6 показателей до удаления выбросов

Поиск и удаление выбросов осуществлялось посредством использования скрипта:

Построение диаграммы размаха:

```
boxplot(Base_10[,3:12])
```

Определение числовых значений выбросов:

```
for (k in Base_10[3:12]){
  print(boxplot.stats(k)$out)}
```

Определение Координат выбросов:

```
for (t in Base_10[3:12]){
  print(which( t %in% boxplot.stats(t)$out))}
```

Удаление выбросов:

```
Base_E[c(53, 56, 69, 102, 104, 170, 178, 209),29] <- NA
```

Диаграмма размахов для 6 показателей после удаления выбросов представлена на рис. 3.3

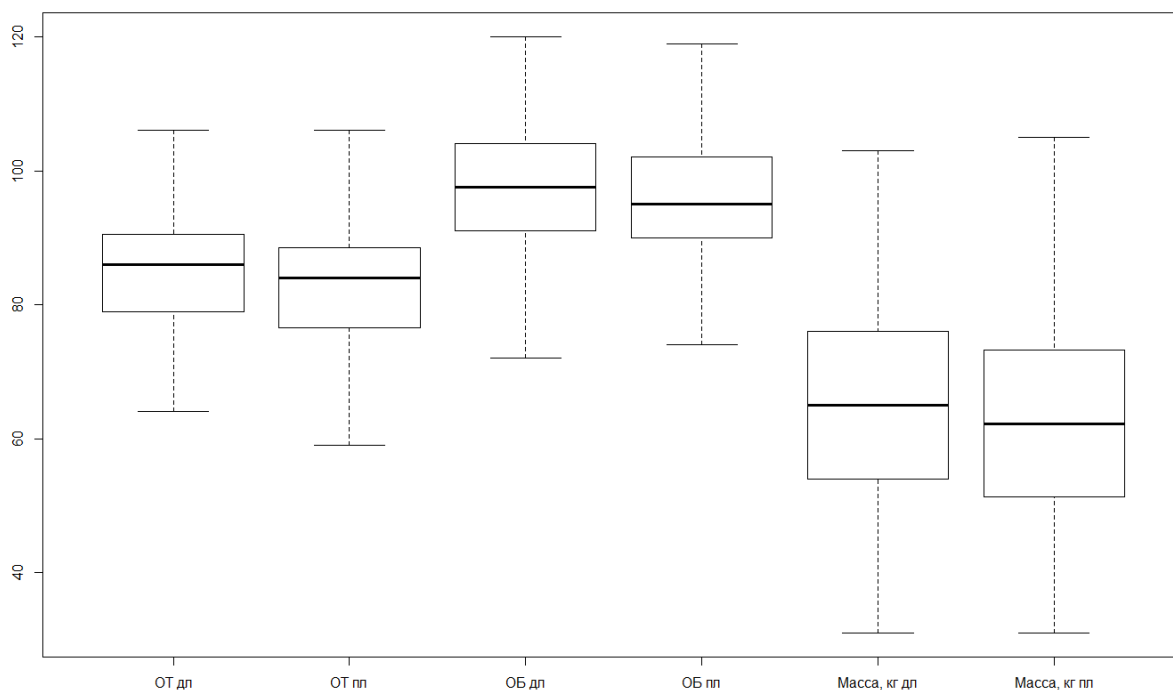


Рисунок 3.3 Диаграмма размаха для 6 после удаления выбросов

Таблица 3.1 – Наименование и границы нормы исследуемых лабораторных данных

| Наименование показателя | Нижняя граница нормы | Верхняя граница нормы |
|---|----------------------|-----------------------|
| Избыток, % | 5 | 10 |
| Индекс массы тела (ИМТ), усл. ед. | 18 | 24,9 |
| Систолическое артериальное давление (САД), мм.рт.ст. | 110 | 125 |
| Диастолическое артериальное давление (ДАД), мм.рт.ст. | 50 | 67 |
| Гомеостатическая оценка модели (НОМА), ммоль/л | 0 | 4 |
| Толерантность к физической нагрузке (ТФН), усл.ед. | 40 | 100 |
| Триацилглицерин (ТАГ), ммоль/л | 0,15 | 1,71 |
| Холестерин, ммоль/л | 3 | 4,4 |
| Холестерин липопротеидов высокой плотности (аХС), ммоль/л | 1,68 | 2 |
| Липопротеиды низкой плотности (ЛПНП), ммоль/л | 0 | 2,85 |
| Анти-тела к териоиднойпероксидазе (АТ ТПО), МЕ/мл | 0 | 30 |
| Инсулин, % | 0 | 11 |
| Кортизол, % | 200 | 400 |
| В-лимфоциты, % | 9 | 29 |
| Т-лимфоциты, % | 40 | 69 |
| Т-хелперы, % | 23 | 45 |
| Т-супрессор, % | 22 | 35 |
| Циркулирующий иммунный комплекс (ЦИК), усл. ед. | 0 | 90 |
| Глюкоза, ммоль/л | 3,5 | 5,7 |
| Каталаза, мккатал/л | 0 | 30 |

| | | |
|-------------------------------------|------|-----|
| Сиаловые кислоты, ммоль/л | 1,9 | 2,5 |
| Церулоплазмин, мг/л | 280 | 400 |
| Лизоцим, % | 45 | 90 |
| Иммуноглобулин класса А (IgA), г/л | 1,25 | 2,8 |
| Иммуноглобулин класса М (IgM), г/л | 1,03 | 2,2 |
| Иммуноглобулин класса G (IgG), г/л | 8,4 | 17 |
| Общие липиды, г/л | 4 | 8 |
| Малоновый диальдегид (МДА), ммоль/л | 0 | 3,8 |
| Тиреотропный гормон (ТТГ), мМЕ/л | 0,3 | 4 |
| Трийодтиронин (Т3), нмоль/л | 1 | 2,8 |
| Тироксин (Т4), нмоль/л | 40 | 120 |
| Интерлейкин-4 (ИЛ-4), пг/мл | 0 | 20 |
| Фактор некроза опухоли (ФНО), пг/мл | 0 | 2,5 |
| Интерлейкин-1В (ИЛ-1В), пкг/мл | 0 | 50 |
| Иммунофермент (ИНФ), пг/мл | 0 | 45 |

После того как выбросы были удалены получена полностью очищенная от выбросов база, которую можно использовать для дальнейшего анализа и восстановления данных.

3.3 Визуальное исследование структуры пропущенных данных

Большое количество статистических методов предполагает, что в ходе наблюдений были получены полностью укомплектованные наборы данных. Но на практике пропуски в данных все же являются повсеместным явлением, прежде чем перейти к анализу данных и построению визуальных образов, необходимо привести обрабатываемый набор данных к "каноническому" виду, т.е. либо удалить фрагменты объектов с недостающими элементами, либо заменить имеющиеся пропуски на некоторые разумные значения. Процедура "борьбы с пропусками" обычно включает следующие шаги:

1. Идентификация недостающих данных.
2. Исследование закономерностей появления отсутствующих значений.
3. Формирование наборов данных, не содержащих пропуски (в результате удаления или замены соответствующих фрагментов).

3.3.1 Идентификация недостающих данных

Идентификация недостающих данных является единственным однозначным шагом. Анализ, пропущенных данных, зависит от понимания процессов, которые воспроизводят экспериментальную информацию. Решение о способе устранения пропущенных значений также будет зависеть от оценки того, какие процедуры приведут к самым надежным и точным результатам. Идентификация пропущенных значений обычно делается с использованием функций `is.na()` и `complete.cases()`:

Список строк, в которых хотя бы одно пропущенное значение:

```
# A tibble: 276 x 108
  Ф.И.О. группа ОТ_дл ОТ_пл ОБ_дл ОБ_пл Масса_дл Масса_пл Избыток_дл Избыток_пл ИМТ_дл ИМТ_пл ТМТ_дл
  <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 person 1     1  NA   NA   NA   NA   46.0  42.0  35.3  23.5  24.2  22.0  NA
2 person 11    1  NA   NA   76   75   66.5  62.0  20.9  12.7  24.5  22.9  43.7
3 person 12    1  NA   NA   NA   NA   54.0  52.0  68.0  62.0  28.4  27.4  35.5
4 person 13    1  NA   NA   NA   NA    NA    NA    NA    NA    NA    NA    NA
5 person 14    1  NA   NA   NA   NA   50.5  49.0  20.2  16.7  24.0  23.3  34.3
6 person 15    1  NA   NA   NA   NA   49.0  47.0  69.0  62.0  28.8  27.6  NA
7 person 16    1  NA   NA   NA   NA   73.0  67.0  52.0  39.6  29.2  26.8  32.9
8 person 41    1  NA   NA   NA   NA    NA    NA    NA    NA    NA    NA    NA
9 person 42    1  NA   NA   NA   NA   50.0  45.5  61.0  46.8  26.3  23.9  138.0
10 person 43   1  NA   NA   NA   NA   60.0  55.0  58.0  45.0  30.0  27.5  142.5
# ... with 266 more rows, and 95 more variables: ТМТ_пл <dbl>, САД_дл <dbl>, САД_пл <dbl>, ДАД_дл <dbl>,
```

Рисунок 3.4 Пропущенные значения

Таким образом в исходном наборе данных 0 полностью укомплектованных строк и 276 строк, имеющих хотя бы одно пропущенное значение. В таблице 3.2 представлено количество пропусков по каждому показателю.

Таблица 3.2 Пропуски

| Показатель | дл/пл | Количество пропусков |
|------------------|-------|----------------------|
| ОБ | дл | 64 |
| | пл | 65 |
| Масса | дл | 33 |
| | пл | 33 |
| Избыток, % | дл | 41 |
| | пл | 43 |
| ИМТ | дл | 37 |
| | пл | 40 |
| ТМТ | дл | 70 |
| | пл | 60 |
| САД, мм. рт. ст. | дл | 29 |
| | пл | 32 |
| ДАД, мм. рт. ст. | дл | 32 |

| | | |
|----------------|----|-----|
| | пл | 36 |
| НОМА | дл | 80 |
| | пл | 95 |
| ТФН | дл | 52 |
| | пл | 62 |
| Общие липиды | дл | 53 |
| | пл | 69 |
| ТАГ | дл | 90 |
| | пл | 89 |
| Холестерин | дл | 45 |
| | пл | 67 |
| аХС | дл | 65 |
| | пл | 77 |
| ЛПОНП | дл | 93 |
| | пл | 89 |
| ЛПНП | дл | 60 |
| | пл | 75 |
| Глюкоза | дл | 45 |
| | пл | 67 |
| Каталаза | дл | 100 |
| | пл | 110 |
| IgA | дл | 56 |
| | пл | 95 |
| IgG | дл | 52 |
| | пл | 86 |
| IgM | дл | 60 |
| | пл | 93 |
| ЦИК | дл | 55 |
| | пл | 92 |
| Т-лим | дл | 74 |
| | пл | 90 |
| Тхел | дл | 58 |
| | пл | 95 |
| Т-суп | дл | 54 |
| | пл | 93 |
| В-лим | дл | 55 |
| | пл | 98 |
| Т4 (53-158) | дл | 123 |
| | пл | 140 |
| Т4св (10-23.2) | дл | 79 |
| | пл | 78 |
| ТТГ (0,23-3,4) | дл | 33 |
| | пл | 56 |
| Инсулин | дл | 68 |
| | пл | 80 |
| Кортизол | дл | 69 |
| | пл | 82 |
| Лептин | дл | 79 |
| | пл | 92 |
| ИЛ-6 | дл | 139 |

| | | |
|---------------------|----|-----|
| | пл | 61 |
| ФНО (не >2,5 пг/мл) | дл | 58 |
| | пл | 140 |
| Гаркави | дл | 85 |
| | пл | 90 |
| КК | дл | 66 |
| | пл | 74 |
| ПИ | дл | 72 |
| | пл | 70 |
| МГ | дл | 68 |
| | пл | 72 |
| АПФ | дл | 64 |
| | пл | 73 |
| NO | дл | 77 |
| | пл | 80 |
| ок ЛПНП | дл | 95 |
| | пл | 99 |

Дополнительную статистическую информацию о пропусках предоставляет функция `md.pattern()`, а функции `aggr()`, `matrixplot()` и `scattMiss()` позволяют отобразить на графиках закономерности во встречаемости отсутствующих наблюдений:

Функция `aggr()` графически отображает число наблюдений для каждой отдельной переменной и для каждой комбинации переменных.

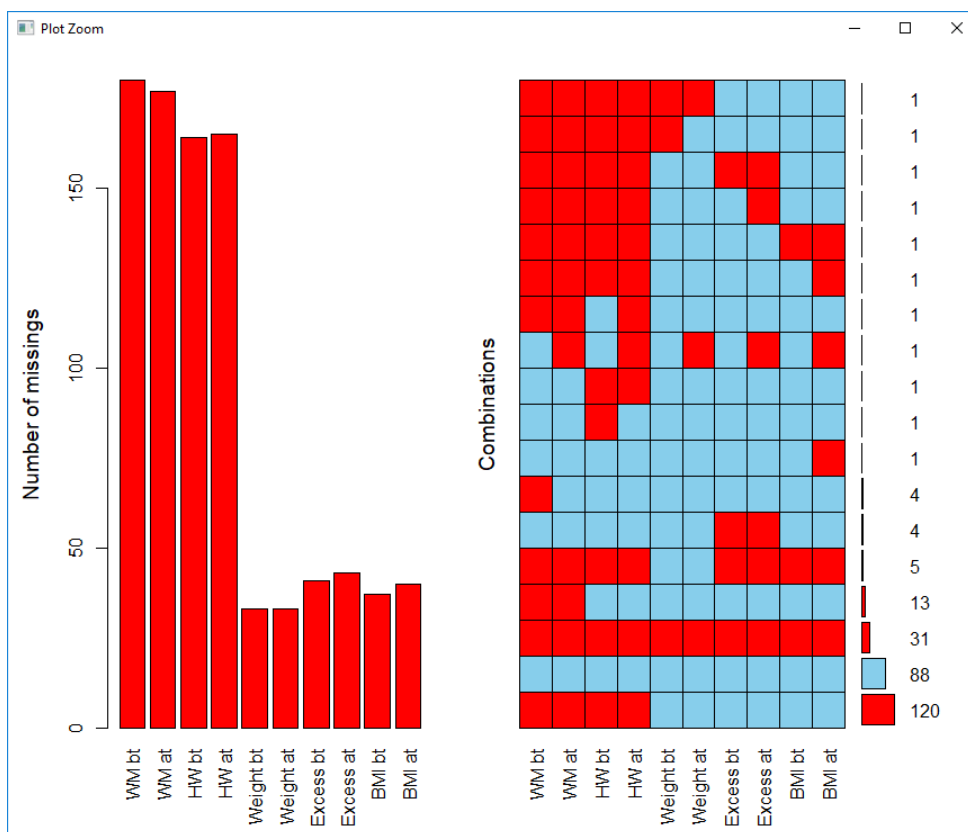


Рисунок 3.5 Результат работы функции `aggr()`

Из фрагмента, представленного на рис. 3.5 видно, что больше всего пропущено значений содержится в показателе ОТ дл.

3.3.2 Исследование закономерностей появления отсутствующих значений

Для анализа закономерностей появления отсутствующих значений можно заменить данные условными значениями: 1 – обозначает пропущенное значение, 0 – имеющееся. Полученную таблицу иногда называют матрицей теней (*shadow matrix*). Вычисление корреляций между этими преобразованными переменными и между исходными переменными поможет узнать, значения каких переменных имеют тенденцию отсутствовать согласованно, а также выявить связи между отсутствием значений в одной переменной и значениями других переменных.

Элементы таблицы данных x равны 1, если соответствующий элемент набора данных отсутствует, и 0 – в противном случае. В этом можно убедиться, рассмотрев первые несколько строк каждой таблицы:

```
x <- as.data.frame(abs(is.na(Base [,3:12])))
```

| ОТ_дл | ОТ_пл | ОБ_дл | ОБ_пл | Масса_дл | Масса_пл | Избыток_дл | Избыток_пл | ИМТ_дл | ИМТ_пл |
|---|-------|-------|-------|----------|----------|------------|------------|--------|--------|
| NA | NA | NA | NA | 46.0 | 42.0 | 35.3 | 23.5 | 24.2 | 22.0 |
| NA | NA | 76 | 75 | 66.5 | 62.0 | 20.9 | 12.7 | 24.5 | 22.9 |
| NA | NA | NA | NA | 54.0 | 52.0 | 68.0 | 62.0 | 28.4 | 27.4 |
| NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| NA | NA | NA | NA | 50.5 | 49.0 | 20.2 | 16.7 | 24.0 | 23.3 |
| NA | NA | NA | NA | 49.0 | 47.0 | 69.0 | 62.0 | 28.8 | 27.6 |
| NA | NA | NA | NA | 73.0 | 67.0 | 52.0 | 39.6 | 29.2 | 26.8 |
| NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| NA | NA | NA | NA | 50.0 | 45.5 | 61.0 | 46.8 | 26.3 | 23.9 |
| NA | NA | NA | NA | 60.0 | 55.0 | 58.0 | 45.0 | 30.0 | 27.5 |
| # ... with 266 more rows, and 95 more variables | | | | | | | | | |
| ОТ_дл | ОТ_пл | ОБ_дл | ОБ_пл | Масса_дл | Масса_пл | Избыток_дл | Избыток_пл | ИМТ_дл | ИМТ_пл |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| # ... with 266 more rows, and 95 more variables | | | | | | | | | |

Рисунок 3.6 Результат замены

Извлекаем переменные, в которых пропущено несколько значений:

```
y <- x[, which(colSums(x) > 0)]
```

Вычисляем корреляцию между этими преобразованными переменными:

```
print(cor(y),4)
```

| | ОТ_дл | ОТ_пл | ОБ_дл | ОБ_пл | Масса_дл | Масса_пл | Избыток_дл | Избыток_пл | ИМТ_дл | ИМТ_пл |
|------------|--------|--------|--------|--------|----------|----------|------------|------------|--------|--------|
| ОТ_дл | 1.0000 | 0.9606 | 0.8527 | 0.8594 | 0.2691 | 0.2457 | 0.2195 | 0.2088 | 0.2873 | 0.2574 |
| ОТ_пл | 0.9606 | 1.0000 | 0.8742 | 0.8964 | 0.2756 | 0.2756 | 0.2274 | 0.2380 | 0.2943 | 0.2864 |
| ОБ_дл | 0.8527 | 0.8742 | 1.0000 | 0.9775 | 0.3045 | 0.2818 | 0.2622 | 0.2533 | 0.3252 | 0.2983 |
| ОБ_пл | 0.8594 | 0.8964 | 0.9775 | 1.0000 | 0.3023 | 0.3023 | 0.2595 | 0.2709 | 0.3227 | 0.3167 |
| Масса_дл | 0.2691 | 0.2756 | 0.3045 | 0.3023 | 1.0000 | 0.9656 | 0.8195 | 0.7962 | 0.8710 | 0.8317 |
| Масса_пл | 0.2457 | 0.2756 | 0.2818 | 0.3023 | 0.9656 | 1.0000 | 0.8195 | 0.8270 | 0.8710 | 0.8634 |
| Избыток_дл | 0.2195 | 0.2274 | 0.2622 | 0.2595 | 0.8195 | 0.8195 | 1.0000 | 0.9723 | 0.9121 | 0.8699 |
| Избыток_пл | 0.2088 | 0.2380 | 0.2533 | 0.2709 | 0.7962 | 0.8270 | 0.9723 | 1.0000 | 0.8866 | 0.8732 |
| ИМТ_дл | 0.2873 | 0.2943 | 0.3252 | 0.3227 | 0.8710 | 0.8710 | 0.9121 | 0.8866 | 1.0000 | 0.9557 |
| ИМТ_пл | 0.2574 | 0.2864 | 0.2983 | 0.3167 | 0.8317 | 0.8634 | 0.8699 | 0.8732 | 0.9557 | 1.0000 |

Рисунок 3.7 Корреляция

Корреляционная матрица между частотами пропусков показывает, что появление недостающих значений для ОТ дл и ОТ пл взаимосвязано ($r = 0.960630$).

Связь между наличием пропущенных значений одной переменной и наблюдаемыми значениями других переменных представлена на рис. 3.8.

```
print(cor(Base_10_1[,3:12], y, use = "pairwise.complete.obs"),4)
```

| | ОТ_дл | ОТ_пл | ОБ_дл | ОБ_пл | Масса_дл | Масса_пл | Избыток_дл | Избыток_пл | ИМТ_дл | ИМТ_пл |
|------------|------------|----------|-----------|-----------|-----------|----------|------------|------------|---------|----------|
| ОТ_дл | NA | -0.11753 | 0.117917 | -0.149275 | NA | -0.11753 | 0.159014 | 0.08930 | NA | -0.06912 |
| ОТ_пл | -0.2747252 | NA | 0.111357 | -0.104732 | NA | NA | 0.176036 | 0.17604 | NA | -0.02707 |
| ОБ_дл | -0.3448796 | -0.43320 | NA | -0.263413 | NA | -0.15499 | 0.014088 | -0.05794 | NA | -0.01020 |
| ОБ_пл | -0.3127647 | -0.36915 | 0.191117 | NA | NA | NA | 0.032554 | 0.03255 | NA | 0.12662 |
| Масса_дл | 0.0185576 | -0.03630 | -0.017968 | -0.042957 | NA | -0.10869 | 0.007922 | -0.04913 | 0.03890 | 0.07352 |
| Масса_пл | -0.0104516 | -0.04865 | -0.043105 | -0.051851 | 0.179771 | NA | 0.009102 | -0.01584 | 0.03787 | 0.11707 |
| Избыток_дл | 0.2374293 | 0.22625 | 0.259880 | 0.252387 | -0.002218 | 0.04994 | NA | 0.14358 | 0.12557 | 0.14669 |
| Избыток_пл | 0.1843889 | 0.16973 | 0.208365 | 0.198527 | -0.003195 | 0.04089 | NA | NA | 0.12462 | 0.15768 |
| ИМТ_дл | 0.0589652 | 0.01706 | 0.064553 | 0.041942 | 0.180215 | 0.07500 | 0.167625 | 0.14050 | NA | 0.09606 |
| ИМТ_пл | -0.0008805 | -0.03633 | 0.004134 | -0.008349 | 0.188535 | 0.17344 | 0.184298 | 0.19441 | NA | NA |

warning message:
In cor(Base_w_e[, 3:12], y, use = "pairwise.complete.obs") :
стандартное отклонение нулевое

Рисунок 3.8 Связь между наличием пропущенных значений

Строки – это наблюдаемые переменные, столбцы – это преобразованные переменные, несущие информацию о наличии пропущенных значений.

В первом столбце этой корреляционной матрицы можно заметить, что частота пропусков показателя (ОТ дл) с большей вероятностью связана с показателем (Избыток дл = 0,237), показателем (Избыток пл = 0,184) и показателем (ИМТ дл = 0, 059).

Анализ пропусков в исходных данных показал, что процент пропусков не превышает 15 в некоторых случаях 20 процентов. Следовательно, можно переходить к процессу восстановления.

3.4 Метод множественного восстановления пропущенных данных

Идея множественного восстановления пропущенных данных хорошо реализована в таком пакете R, как `mice`.

Функция `mice()` (многомерное восстановление данных при помощи связанных уравнений – *multivariate imputation by chained equations*) использует исходную таблицу данных с пропущенными значениями, а возвращает объект, содержащий несколько полных наборов данных (пять по умолчанию). Каждый такой полный набор данных получается при восстановлении пропущенных данных исходной таблицы. В алгоритме восстановления данных есть случайная составляющая, поэтому все производные полные наборы данных немного отличаются друг от друга.

Замещение пропущенных значения осуществляется при помощи выборки Гиббса. По умолчанию значения каждой переменной, содержащей пропущенные значения, предсказываются по значениям остальных переменных. Полученные уравнения используются для замещения пропущенных данных подходящими значениями.

Пользователь может выбирать вид предсказательной модели для каждой переменной и определять переменные, которые в нее войдут. По умолчанию для замены пропущенных значений непрерывных переменных используется метод соответствия, предсказанного среднего (*predictive mean matching*), а логистическая или полиномиальная логистическая регрессия применяются для дихотомических (фактор с двумя уровнями) или полиномиальных (фактор с более чем двумя уровнями) переменных соответственно.

Алгоритм работы данного пакета представлен на рис. 3.9:

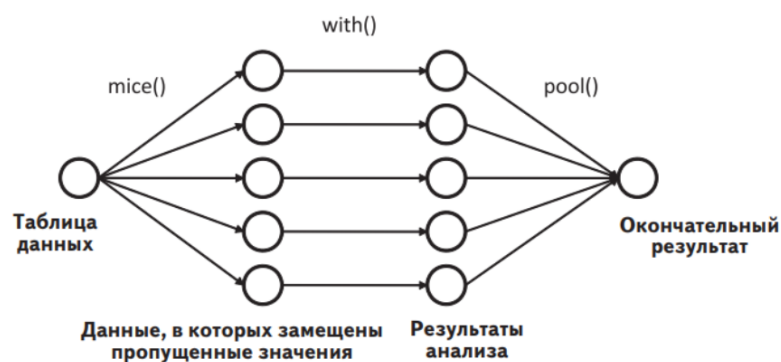


Рисунок 3.9 Алгоритм работы функции `mice`

Процесс заполнения пропусков реализуется в несколько этапов:

1. Функция `mice()` обрабатывает `mydata` – матрицу или структуру данных, содержащую недостающие значения, и возвращает объект `imp`, содержащий несколько полных наборов данных, отличающихся между собой значениями пропусков, заполненных в ходе процедуры Гиббса, но с включением некоторого случайного компонента: `mice(data, m)`;
2. При помощи функция `with()` применяется статистическая модель `analysis` к каждому набору данных из `imp` (например, линейная `lm()`, обобщенная линейная `glm()` или иная, задаваемая пользователем): `fit <- with(imp, analysis)`;
3. Функция `pool()` объединяет результаты, полученные для отдельных производных наборов данных: `pooled <- pool(fit)`;
4. Функция `complete()` создает набор данных с заполненными пропусками: `complete(imp, action=#)`.

После применения функции `mice ()` анализируемые данные принимают следующую структуру (рис.3.10). Из представленной информации видно, что было создано пять искусственных наборов данных (Number of multiple imputations). К каждой переменной с пропущенными значениями был применен метод сочетания, предсказанного среднего (`ppm, predictive mean matching`). Восстановление данных применялось ко всем показателям базы. Строка `VisitSequence` содержит информацию о последовательности замещения данных в переменных – справа налево, начиная с `OT` дл и заканчивая `окЛПНП` пл. Из матрицы `PredictorMatrix` следует, что при восстановлении пропущенных

данных каждой переменной использовались все остальные переменные. В этой матрице строки соответствуют переменным, значения которых восстанавливались, столбцы – переменным, которые использовались для расчета замещаемых значений, а 1/0 указывает, использовалась ли переменная: да/нет.

```

Multiply imputed data set
Call:
mice(data = Base[3:12])
Number of multiple imputations: 5
Missing cells per column:
      OT дл      OT пл      ОБ дл      ОБ пл  Масса, кг дл  Масса, кг пл  Избыток, % дл  Избыток, % пл
      180      177      164      165      33      33      41      43
      ИМТ дл      ИМТ пл
      37      40
Imputation methods:
      OT дл      OT пл      ОБ дл      ОБ пл  Масса, кг дл  Масса, кг пл  Избыток, % дл  Избыток, % пл
      "pmm"      "pmm"      "pmm"      "pmm"      "pmm"      "pmm"      "pmm"      "pmm"
      ИМТ дл      ИМТ пл
      "pmm"      "pmm"
VisitSequence:
      OT дл      OT пл      ОБ дл      ОБ пл  Масса, кг дл  Масса, кг пл  Избыток, % дл  Избыток, % пл
      1      2      3      4      5      6      7      8
      ИМТ дл      ИМТ пл
      9      10
PredictorMatrix:
      OT дл OT пл ОБ дл ОБ пл Масса, кг дл Масса, кг пл Избыток, % дл Избыток, % пл ИМТ дл ИМТ пл
      OT дл 0 1 1 1 1 1 1 1 1 1
      OT пл 1 0 1 1 1 1 1 1 1 1
      ОБ дл 1 1 0 1 1 1 1 1 1 1
      ОБ пл 1 1 1 0 1 1 1 1 1 1
      Масса, кг дл 1 1 1 1 0 1 1 1 1 1
      Масса, кг пл 1 1 1 1 1 0 1 1 1 1
      Избыток, % дл 1 1 1 1 1 1 0 1 1 1
      Избыток, % пл 1 1 1 1 1 1 1 0 1 1
      ИМТ дл 1 1 1 1 1 1 1 1 0 1
      ИМТ пл 1 1 1 1 1 1 1 1 1 0
Random generator seed value: NA

```

Рисунок 3.10 Структура анализируемых данных при использовании функции mice

Результатом работы функции mice () стала база с восстановленными значениями.

Для оценки точности восстановленных данных был построен график densityplot(imp) рис. 3.11. Плотность восстановленных данных для каждого переменного набора показана красным цветом, а плотность наблюдаемых показателей показана синим цветом. Исходя из этого можно сделать вывод что распределения восстановленных данных совпадают с исходными.

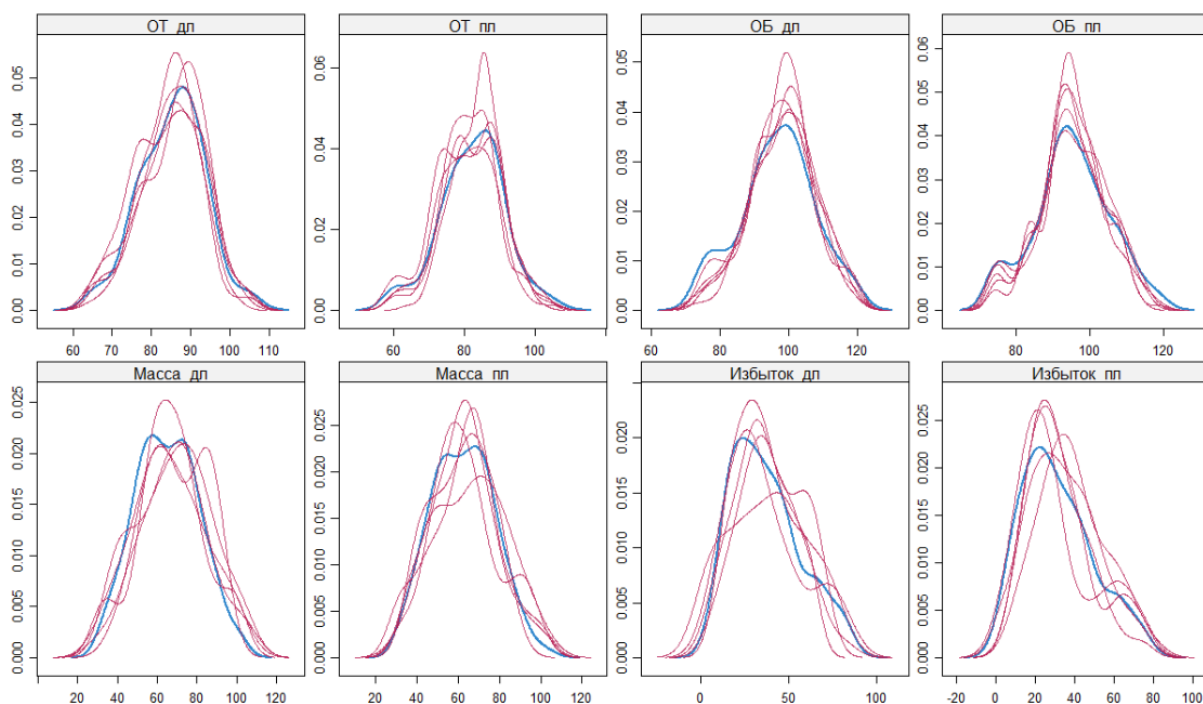


Рисунок 3.11 Плотность временных данных

3.5 Восстановление данных с использованием метода "bagImpute"

Воспользуемся еще одним методом восстановления пропусков. Метод "bagImpute" осуществляет для каждой из имеющихся переменных построение множественной бутстреп-агрегированной модели или бэггинг-модели на основе деревьев регрессии, принимая все остальные переменные в качестве предикторов. Этот метод точен, но требует значительных затрат времени на вычисление, особенно при работе с данными большого объема. Из-за этого было принято решение не использовать его в качестве основного. Результат восстановления пропущенных данные с помощью метода "bagImpute".

Можно оценить, насколько расходятся между собой результаты, полученные каждым способом заполнения. Для этого сформируем блок данных из $3 * 16 = 48$ строк исходной таблицы с заполненными пропусками двумя методами ("mice", "Bag") и выполним редукцию переменных методом главных компонент из многомерного пространства в двухмерное.

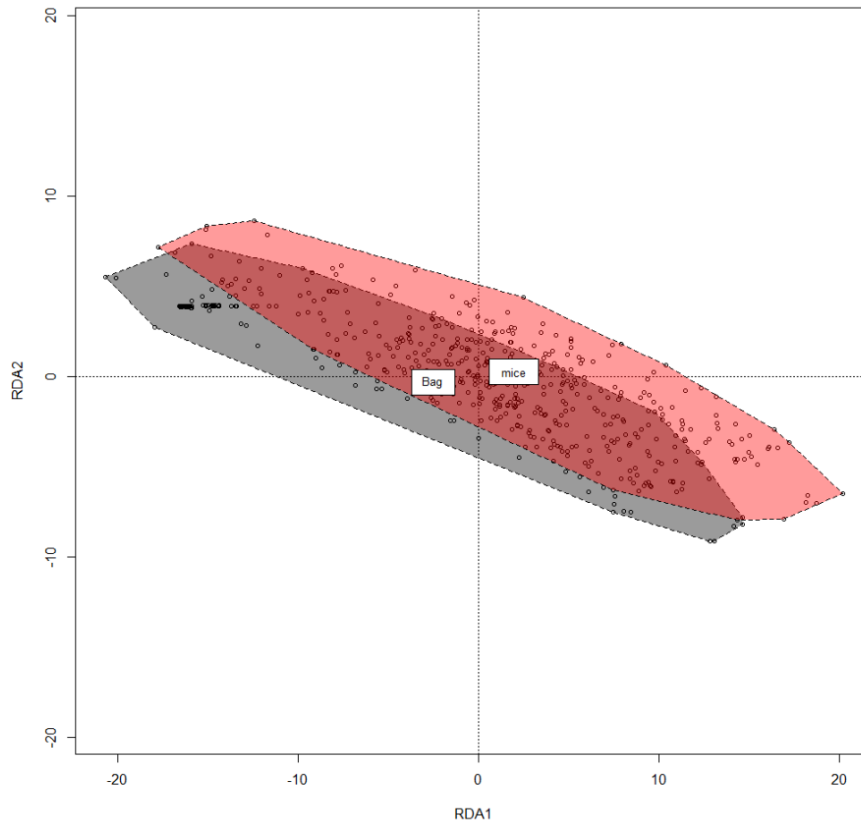


Рисунок 3.12 Графическое представление результатов "mice" и "Bag"

Контурсы, проведенные через крайние точки, области каждого из двух блоков данных метки методов размещены в центрах тяжести полученных многоугольников. Оба метода, учитывающих внутреннюю структуру данных, дали приблизительно похожие результаты.

3.6 Определение информативных признаков

После восстановления пропусков в исходных наборах были получены полностью укомплектованный набор данных, который можно использовать для моделирования визуальных образов используя лица Чернова. Но так как количество показателей намного больше чем можно поместить в лица, необходимо подготовить восстановленные данные к построению визуальных образов. Подготовка данных предполагает определения наиболее информативных признаков из всего восстановленного набора данных. Расчет информативности осуществлялся с использованием метода Кульбака.

После расчета информативности по 5 группам были получены следующие результаты, которые представлены в таблице 3.3.

Таблица 3.3 – Информативность признаков по пяти группам

| Группа 1 | | Группа 2 | | Группа 3 | | Группа 4 | | Группа 5 | |
|----------|------|----------|------|----------|------|----------|------|----------|------|
| ТМТ | 33,3 | ТМТ | 30,4 | ТМТ | 26,6 | Каталаза | 19,4 | НО | 34 |
| Гаркави | 14,2 | НО | 19,1 | НО | 15,3 | НО | 18 | Каталаза | 21,7 |
| НО | 11,4 | Гаркави | 11 | Каталаза | 15,3 | ТМТ | 17,5 | Гаркави | 12,3 |
| Каталаза | 7,9 | Лептин | 7,9 | Гаркави | 14,6 | Гаркави | 8,9 | ТМТ | 9,6 |
| ФНО | 7,8 | Каталаза | 5,9 | Лептин | 12,5 | АПФ | 4,5 | АПФ | 8,1 |
| Лептин | 6,2 | АПФ | 4,4 | ФНО | 10,5 | ФНО | 4,4 | ФНО | 5,8 |
| КК | 2,9 | ФНО | 4 | АПФ | 5,8 | ИЛ-4 | 2,9 | ИЛ-4 | 5,2 |
| АПФ | 2,5 | ИЛ-4 | 4 | окЛПНП | 5,5 | Лептин | 2,3 | КК | 4,7 |
| окЛПНП | 2,2 | КК | 1 | КК | 5,2 | окЛПНП | 1,9 | Лептин | 2,8 |
| ИЛ-4 | 2,2 | окЛПНП | 0,7 | ИЛ-4 | 3 | КК | 1 | окЛПНП | 0,8 |

3.7 Построение визуальных образов

Для построения лиц Чернова использована библиотека TeachingDemos функция faces. Набор конструктивных параметров лиц Чернова представлен в таблице 3.4.

Таблица 3.4 – Конструктивные параметры


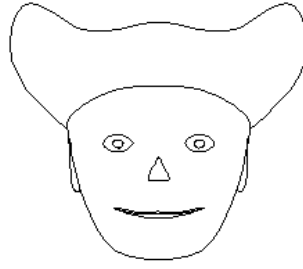
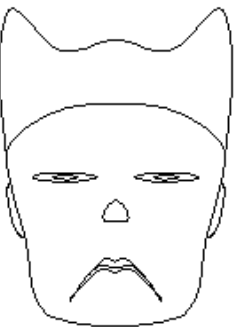



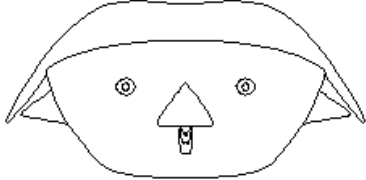

| Конструктивные параметры лица | Информативные признаки |
|-------------------------------|------------------------|
| height of face – высота лица | ТМТ |
| width of face – ширина лица | Гаркави |
| shape of face – форма лица | НО |
| height of mouth – высота рта | Каталаза |
| width of mouth – ширина рта | ФНО |
| curve of smile – кривизна рта | Лептин |
| height of eyes – высота глаз | АПФ |
| width of eyes – ширина глаз | ИЛ-4 |
| height of hair – рост волос | окЛПНП |
| width of hair – ширина волос | КК |



Для построения лиц Чернова был использован следующий скрипт:

```
require("TeachingDemos")
faces(bt_10[,2:11],labels = bt_10$Группа)#до лечения
title("До лечения")
faces(at_10[,2:11],labels = at_10$Группа)#после лечения
title("После лечения")
```


Построенные визуальные образы позволяют сравнить результаты до и после лечения (таблица 3.5).

Таблица 3.5 – Визуальные образы до и после лечения

| № группы | До лечения | После лечения |
|----------|---|---|
| Группа 1 |  |  |
| Группа 2 |  |  |
| Группа 3 |  |  |
| Группа 4 |  |  |

| | | |
|----------|---|---|
| Группа 5 |  |  |
|----------|---|---|

Анализируя результаты можно заметить по каким группам лечение было наиболее эффективным. Можно выделить группу 1, по этой группе заметны значительные изменения. Визуальные образы группы 2 и группы 5 практически не изменились, следовательно, можно сделать вывод, что лечение не оказало положительного результата.

Глава 4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Целью раздела является проектирование и создание конкурентоспособных разработок, технологий, отвечающих современным требованиям в области ресурсоэффективности и ресурсосбережения.

Достижение цели обеспечивается решением задач:

- Оценка коммерческого потенциала и перспективности проведения научных исследований;
- Определение возможных альтернатив проведения научных исследований, отвечающих современным требованиям в области ресурсоэффективности и ресурсосбережения;
- Планирование научно-исследовательских работ;
- Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования.

В работе будет рассматриваться информационная технология анализа результатов лечения детей с различными формами ожирения. Данная технология предназначена для анализа пропущенных данных, восстановления пропусков в исходных наборах данных и комплексного анализа результатов лечения с применением методов визуализации. Данная разработка нацелена на использование внутри НИИ курортологии и физиотерапии г. Томска.

4.1. Оценка коммерческого потенциала и перспективности проведения научных исследований

4.1.1. Потенциальные потребители результатов исследования

Для анализа потребителей результатов исследования необходимо рассмотреть целевой рынок и провести его сегментирование.

Целевым рынком проекта являются медицинские учреждения (заказчиком является НИИ курортологии и физиотерапии г. Томска). Критерии сегментирования данной работы – профессия в медицинской сфере,

потребители, социальная принадлежность, ступени образования, научная степень, звание, вид потребителей, класс информационных технологий. Выделим наиболее значимые для разрабатываемой информационной технологии критерии: класс медицинской информационной технологии и потребители. На основе выявленных данных построим карту сегментирования, она представлена на рисунке 4.1.

| | | Класс медицинских информационных технологий | | |
|-------------|--|--|---|---|
| | | Технологии комплексной компьютеризации медицинских организаций и подразделений | Информационные технологии комплексного анализа медицинских данных | Информационные технологии федерального, регионального и муниципального уровней, управление здравоохранением, состояние здоровья населения |
| Потребители | Амбулаторно-поликлинические учреждения | | | |
| | Медицинские научно-исследовательские учреждения, НИИ | | | |
| | Диагностические центры, стационары, клиники | | | |

Рисунок 4.1 Карта сегментирования рынка услуг:

| | | | |
|--|--------------------------------|--|--|
| | Информационная система «ЕГИСЗ» | | Информационная система «Анализ данных» |
|--|--------------------------------|--|--|

В результате построения карты сегментирования выявлено, какие ниши на рынке услуг не заняты конкурентами или где уровень конкуренции низок.

Исходя из полученных данных, было принято решение разработать информационную технологию для комплексного анализа медицинских данных с применением визуальных образов основным потребителем которой будут медицинские научно-исследовательские учреждения, НИИ.

4.1.2. Анализ конкурентных технических решений

Проведем сравнительный анализ конкурентных технических решений, существующих на рынке: «ЕГИСЗ», «Анализ данных». Анализируемые данные представлены в таблице 4.1.

Таблица 4.1 – Оценочная карта для сравнения конкурентных технических решений (разработок)

| Критерии оценки | Вес критерия | Баллы | | | Конкурентоспособность | | |
|--|--------------|----------------|-----------------|-----------------|-----------------------|-----------------|-----------------|
| | | Б _ф | Б _{к1} | Б _{к2} | К _ф | К _{к1} | К _{к2} |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Технические критерии оценки ресурсоэффективности | | | | | | | |
| 1. Повышение производительности труда пользователя | 0,13 | 5 | 5 | 4 | 0,65 | 0,65 | 0,52 |
| 2. Удобство в эксплуатации | 0,05 | 5 | 4 | 5 | 0,25 | 0,2 | 0,25 |
| 3. Надежность | 0,15 | 4 | 5 | 4 | 0,6 | 0,75 | 0,6 |
| 4. Потребность в ресурсах памяти | 0,04 | 5 | 4 | 5 | 0,2 | 0,16 | 0,2 |
| 5. Безопасность | 0,07 | 5 | 5 | 5 | 0,35 | 0,35 | 0,35 |
| 6. Функциональная мощность (предоставляемые возможности) | 0,08 | 4 | 3 | 4 | 0,32 | 0,24 | 0,32 |
| 7. Простота эксплуатации | 0,06 | 5 | 5 | 4 | 0,3 | 0,3 | 0,24 |
| 8. Качество интеллектуального интерфейса | 0,1 | 5 | 4 | 4 | 0,5 | 0,4 | 0,4 |
| Экономические критерии оценки эффективности | | | | | | | |
| 1. Конкурентоспособность продукта | 0,1 | 5 | 3 | 4 | 0,5 | 0,3 | 0,4 |
| 2. Цена | 0,15 | 5 | 2 | 3 | 0,75 | 0,3 | 0,45 |
| 3. Срок выхода на рынок | 0,01 | 4 | 5 | 5 | 0,04 | 0,05 | 0,05 |
| 4. Предполагаемый срок эксплуатации | 0,01 | 5 | 3 | 4 | 0,05 | 0,03 | 0,04 |
| 5. Послепродажное обслуживание | 0,01 | 5 | 4 | 3 | 0,05 | 0,04 | 0,03 |
| 6. Финансирование научной разработки | 0,04 | 5 | 2 | 2 | 0,2 | 0,08 | 0,08 |
| Итого | 1 | 67 | 54 | 56 | 4,76 | 3,85 | 3,93 |

Анализ конкурентных технических решений определяется по формуле 4.1:

$$K = \sum B_i \cdot B_i, \quad (4.1)$$

где K – конкурентоспособность научной разработки или конкурента;

B_i – вес показателя (в долях единицы);

B_i – балл i -го показателя.

Основываясь на знаниях о конкурентах, обратим внимание на сильные стороны разрабатываемой информационной технологии:

- Низкая цена (в три раза ниже, чем у конкурентов).
- Длительный срок эксплуатации (5 лет гарантии, по сравнению с годом гарантии у конкурентов).
- Качество послепродажного обслуживания (поддержка продукта в течение всего срока гарантии).
- Для разработки информационной технологии требовались минимальные вложения.

4.1.3. Технология QuaD

Для анализа разрабатываемого проекта применена технология QuaD. Данная технология позволяет измерить характеристики качества разработки и ее перспективность использования в работе. В основе технологии QuaD лежит нахождение средневзвешенной величины двух групп показателей: качества и потенциала разработки. Каждый показатель оценивается экспертным путем по сто балльной шкале. В таблице 4.2 представлена оценочная карта.

Таблица 4.2 – Оценочная карта для сравнения конкурентных технических решений

| Критерии оценки | Вес критерия | Баллы | Максимальный балл | Относительное значение | Средне взвешенное значение |
|--|--------------|-------|-------------------|------------------------|----------------------------|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Показатели оценки качества разработки | | | | | |
| Надежность | 0,2 | 80 | 100 | 0,8 | 0,16 |
| Унифицированность | 0,1 | 60 | 100 | 0,6 | 0,06 |
| Функциональная мощность | 0,2 | 90 | 100 | 0,9 | 0,18 |
| Простота эксплуатации | 0,2 | 90 | 100 | 0,9 | 0,18 |

| | | | | | |
|---|----------|----|------------|-----|-------------|
| Качество интеллектуального интерфейса | 0,2 | 80 | 100 | 0,8 | 0,16 |
| Показатели оценки коммерческого потенциала разработки | | | | | |
| Конкурентоспособность продукта по сравнению с существующими системами | 0,1 | 70 | 100 | 0,7 | 0,07 |
| Итого: | 1 | | 100 | | 0,81 |

Оценка качества и перспективности по технологии QuaD определяется по формуле 4.2:

$$P_{cp} = \sum B_i \cdot B_i, \quad (4.2)$$

Где P_{cp} – средневзвешенное значение показателя качества и перспективности научной разработки;

B_i – вес показателя (в долях единицы);

B_i – средневзвешенное значение i -го показателя.

Получаем, что $P_{cp} = 0,81$.

Значение P_{cp} позволяет говорить о перспективах разработки и качестве проведенного исследования. Так как $P_{cp} = 0.81$, можно утверждать, что данная разработка является перспективной.

4.1.4. SWOT-анализ

SWOT-анализ является одним из самых распространенных методов анализа и стратегического планирования в менеджменте и маркетинге. Он дает четкое представление о факторах внешней и внутренней среды и указывает, в каких направлениях нужно действовать, используя сильные стороны, чтобы максимизировать возможности и свести к минимуму угрозы и слабые стороны. С помощью этого метода можно обозначить основные проблемы проекта, определить пути решения и перспективу развития.

Объектом для проведения SWOT-анализа является процесс использования информационной технологии для анализа результатов лечения детей с проблемами ожирения.

Результатом анализа является разработка маркетинговой стратегии или гипотезы для дальнейшей проверки, они представлены в таблице 4.3.

Таблица 4.3 – SWOT-анализ

| | Сильные стороны научно-исследовательского проекта: С1. Упрощение работы с медицинскими данными; С2. Сокращение времени обработки медицинских данных; С3. Анализ пропущенных значений; С4. Восстановления пропусков в исходных данных; С5. Комплексная оценка результатов лечения с использованием визуальных образов. | Слабые стороны научно-исследовательского проекта: Сл1. Неквалифицированные пользователи; Сл2. Значительные временные и интеллектуальные затраты на реализацию; Сл3. Текучесть кадров; Сл4. Срок выхода на рынок. |
|--|---|---|
| Возможности: В1. Увеличение скорости проведения анализа медицинских данных; В2. Исключение ошибок, связанных с человеческим фактором; В3. Расширения функционала информационной технологии; В4. Увеличение масштабов проекта. | 1 Быстрый и точный сбор статистики приведет к исключению работы людей с многочисленными данными и позволит получить информативные данные; 2 Решение проблемы отсутствующих данных; 3 Использование визуальных образов сократит время, затраченное на анализ результатов. | 1 Предоставление консультаций клиентам; 2 При разработке учитывать наиболее время затратные этапы, сокращение времени должно быть не на проценты, а в разы сокращать время обработки; 3 Привлечение кадров из ТПУ позволит увеличить темпы работы над проектом. |
| Угрозы: У1. Сбои в работе; У2. Потеря конкурентоспособности; У3. Отказ от технической поддержки проекта после внедрения; У4. Отсутствие спроса на расширение разработки. | 1 Сбои в работе можно исключить за счет грамотного выбора разработчика и предоставления лучших инструментов для работы; 2 Предоставление сервисного обслуживания; 3 Анализ ранка подобных технологий для удовлетворения требований клиентов. | 1 Исследование новых технологий для расширения методов анализа данных; 2 Проводить разработку информационной технологии таким образом, чтобы она была гибкой, легко восприимчивой к изменениям. |

Вывод. В результате SWOT-анализа были рассмотрены сильные и слабые стороны разработки информационной технологии. Особенностью данной информационной технологии является увеличение скорости анализа медицинских данных – это осуществляется посредством применением в работе методов визуализации, основанных на таком типе пиктографиков как лица Чернова.

Решение такой проблемы как пропуски в данных позволит увеличить функциональные возможности информационной технологии посредством включения новых методов обработки и анализа, которые работают только с полными наборами данных.

Согласно этому можно обозначить основные стратегические направления:

1. Проведение консультаций и обучений пользователей с целью повышение квалификации;
2. Расширение функциональных возможностей посредством добавления новых методов анализа данных.

4.2. Планирование научно-исследовательских работ

4.2.1. Структура работ в рамках научного исследования

Планирование комплекса предполагаемых работ осуществляется в следующем порядке:

- Определение структуры работ в рамках научного исследования;
- Определение участников каждой работы;
- Установление продолжительности работ;
- Построение графика проведения научных исследований и разработок.

Для выполнения научного исследования сформирована рабочая группа, в состав которой входят студенты и преподаватель – руководитель проекта. По каждому виду запланированных работ установлена соответствующая должность исполнителей, эти данные представлены в таблице 4.4.

Таблица 4.4 - Перечень этапов, работ и распределение исполнителей

| Основные этапы | № раб. | Содержание работ | Должность исполнителя |
|---|--------|--|-------------------------------|
| Формирование темы и цели разработки | 1 | Продумать идею и записать основные данные | Руководитель проекта, студент |
| Разработка технического задания | 2 | Подробное описание идеи | Руководитель, студент |
| | 3 | Выявление функциональных блоков | Руководитель, студент |
| | 4 | Определения содержания блоков | Руководитель, Студент |
| | 5 | Календарное планирование работ | Руководитель |
| Поиск подходов для решения текущих задач | 6 | Поиск решения по каждому функциональному блоку | Студент |
| | 7 | Реализация найденного решения | Студент |
| | 8 | Тестирование | Студент |
| | 9 | Отладка: устранение ошибок и повторное тестирование блоков проекта | Студент |
| Обобщение и оценка результатов | 10 | Оценка эффективности полученных результатов | Руководитель, студент |
| | 11 | Определение целесообразности проведения ОКР | Руководитель, студент |
| Разработка технической документации и проектирование | 12 | Разработка диаграмм по проекту | Студент |
| | 13 | Выбор проектируемой технологии | Руководитель, студент |
| | 14 | Оценка эффективности информационной технологии | Руководитель, студент |
| Тестирование информационной технологии конечными пользователями | 15 | Тестирование конечными пользователями | Студент |
| | 16 | Доработка | Студент |
| Оформление отчета по НИР (комплекта документации по ОКР) | 17 | Составление пояснительной записки | Руководитель, студент |

4.2.2. Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к.

зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости $t_{ожі}$ используется формула 4.3:

$$t_{ожі} = \frac{3t_{\min i} + 2t_{\max i}}{5}, \quad (4.3)$$

где $t_{ожі}$ – ожидаемая трудоемкость выполнения i -ой работы чел.-дн.;

$t_{\min i}$ – минимально возможная трудоемкость выполнения заданной i -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{\max i}$ – максимально возможная трудоемкость выполнения заданной i -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях T_{pi} , учитывающая параллельность выполнения работ несколькими исполнителями. Такое вычисление необходимо для обоснованного расчета заработной платы, так как удельный вес зарплаты в общей сметной стоимости научных исследований составляет около 65 %. Продолжительно одной работы определяется по формуле 4.4:

$$T_{pi} = \frac{t_{ожі}}{Ч_i}, \quad (4.4)$$

где T_{pi} – продолжительность одной работы, раб. дн.;

$t_{ожі}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Составим таблицу 4.5 и подсчитаем значения $t_{ожі}$ и T_{pi} для каждой из работ.

Таблица 4.5 – Перечень работ и подсчет показателей

| Перечень работ | t min | t max | t ожид | Ч | Tri |
|--|-------|-------|--------|---|-----|
| Продумать идею и записать основные данные | 4 | 7 | 5,2 | 3 | 1,7 |
| Подробное описание идеи | 6 | 8 | 6,8 | 2 | 3,4 |
| Выявление функциональных блоков | 2 | 4 | 2,8 | 2 | 1,4 |
| Определения содержания блоков | 1 | 5 | 2,6 | 2 | 1,3 |
| Календарное планирование работ | 1 | 6 | 3 | 1 | 3 |
| Поиск решения по каждому функциональному блоку | 5 | 10 | 7 | 2 | 3,5 |
| Реализация решений | 50 | 70 | 58 | 2 | 29 |
| Тестирование | 6 | 10 | 7,6 | 1 | 7,6 |
| Отладка: устранение ошибок и повторное тестирование блоков проекта | 5 | 15 | 9 | 2 | 4,5 |
| Оценка эффективности полученных результатов | 6 | 12 | 8,4 | 2 | 4,2 |
| Определение целесообразности проведения ОКР | 3 | 7 | 4,6 | 2 | 2,3 |
| Разработка диаграмм по проекту | 4 | 11 | 6,8 | 1 | 6,8 |
| Выбор проектируемой технологии | 1 | 3 | 1,8 | 2 | 0,9 |
| Оценка эффективности информационной технологии | 3 | 6 | 4,2 | 2 | 2,1 |
| Тестирование технологии | 10 | 15 | 12 | 3 | 4 |
| Исправление ошибок, доработка | 6 | 9 | 7,2 | 1 | 7,2 |
| Написание пояснительной записки | 14 | 20 | 16,4 | 2 | 8,2 |

4.2.3. Разработка графика проведения научного исследования

Составим график проведения научных работ в форме диаграммы Ганта.

Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней переведем в календарные дни. Для этого воспользуемся формулой 4.5:

$$T_{ki} = T_{pi} \cdot k_{кал}, \quad (4.5)$$

где T_{ki} – продолжительность выполнения i -й работы в календарных днях;

T_{pi} – продолжительность выполнения i -й работы в рабочих днях;

$k_{кал}$ – коэффициент календарности.

Коэффициент календарности определяется по формуле 4.6:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}}, \quad (4.6)$$

где $T_{кал}$ – количество календарных дней в году;

$T_{вых}$ – количество выходных дней в году;

$T_{пр}$ – количество праздничных дней в году.

$$k_{кал} = 365 / (365 - 125) = 1,5.$$

Рассчитанные значения в календарных днях по каждой работе T_{ki} округлим до целого числа.

Все рассчитанные значения сведены в таблицу 4.6. Исполнители: Р-руководитель, С-студент. Временные параметры рассчитаны для трех исполнений.

Исполнение 1 – Данный проект.

Исполнение 2 – «ЕГИСЗ»

Исполнение 3 – «Анализ данных».

Таблица 4.6 - Временные показатели проведения научного исследования

| Название работы, исполнители | Трудоёмкость работ | | | | | | | | | Длительность работ в рабочих днях T_{pi} | | | Длительность работ в календарных днях T_{ki} | | |
|---|---------------------|-------|-------|---------------------|-------|-------|--------------------|-------|-------|--|-------|-------|--|-------|-------|
| | t_{min} , чел-дни | | | t_{max} , чел-дни | | | $t_{ож}$, чел-дни | | | | | | | | |
| | Исп.1 | Исп.2 | Исп.3 | Исп.1 | Исп.2 | Исп.3 | Исп.1 | Исп.2 | Исп.3 | Исп.1 | Исп.2 | Исп.3 | Исп.1 | Исп.2 | Исп.3 |
| Продумать идею и записать основные данные. Р,С | 4 | 4 | 5 | 7 | 7 | 8 | 5,2 | 5,2 | 6,2 | 1,7 | 1,7 | 2,1 | 3 | 3 | 3 |
| Подробное описание идеи. Р,С | 6 | 6 | 7 | 8 | 8 | 9 | 6,8 | 6,8 | 7,8 | 3,4 | 3,4 | 3,9 | 5 | 5 | 6 |
| Выявление функциональных блоков. Р,С | 2 | 3 | 3 | 4 | 5 | 5 | 2,8 | 3,8 | 3,8 | 1,4 | 1,9 | 1,9 | 2 | 3 | 3 |
| Определения содержания блоков. Р, С | 1 | 1 | 1 | 5 | 5 | 5 | 2,6 | 2,6 | 2,6 | 1,3 | 1,3 | 1,3 | 2 | 2 | 2 |
| Календарное планирование работ. Р | 1 | 2 | 2 | 6 | 7 | 7 | 3 | 4 | 4 | 3,0 | 4,0 | 4,0 | 5 | 6 | 6 |
| Поиск решения по каждому функциональному блоку. С | 5 | 5 | 5 | 10 | 10 | 10 | 7 | 7 | 7 | 3,5 | 3,5 | 3,5 | 5 | 5 | 5 |
| Реализация найденного решения. С | 50 | 60 | 62 | 70 | 75 | 92 | 58 | 66 | 74 | 29,0 | 33,0 | 37,0 | 44 | 50 | 56 |
| Тестирование. С | 6 | 10 | 10 | 10 | 14 | 14 | 7,6 | 11,6 | 11,6 | 7,6 | 11,6 | 11,6 | 11 | 17 | 17 |

Продолжение таблицы 4.6

| | | | | | | | | | | | | | | | |
|---|------------|------------|------------|------------|------------|------------|--------------|--------------|--------------|-------------|--------------|--------------|------------|------------|------------|
| Отладка: устранение ошибок и повторное тестирование блоков проекта. С | 5 | 4 | 6 | 15 | 14 | 16 | 9 | 8 | 10 | 4,5 | 4,0 | 5,0 | 7 | 6 | 8 |
| Оценка эффективности полученных результатов. Р,С | 6 | 6 | 6 | 12 | 12 | 12 | 8,4 | 8,4 | 8,4 | 4,2 | 4,2 | 4,2 | 6 | 6 | 6 |
| Определение целесообразности проведения ОКР. Р,С | 3 | 4 | 4 | 7 | 8 | 8 | 4,6 | 5,6 | 5,6 | 2,3 | 2,8 | 2,8 | 3 | 4 | 4 |
| Разработка диаграмм по проекту. С | 4 | 4 | 3 | 11 | 11 | 11 | 6,8 | 6,8 | 6,2 | 6,8 | 6,8 | 6,2 | 10 | 10 | 9 |
| Выбор проектируемой технологии. Р,С | 1 | 1 | 1 | 3 | 3 | 3 | 1,8 | 1,8 | 1,8 | 0,9 | 0,9 | 0,9 | 1 | 1 | 1 |
| Оценка эффективности информационной технологии. Р,С | 3 | 3 | 4 | 6 | 6 | 7 | 4,2 | 4,2 | 5,2 | 2,1 | 2,1 | 2,6 | 3 | 3 | 4 |
| Тестирование конечными пользователями. С | 10 | 13 | 14 | 15 | 20 | 23 | 12 | 15,8 | 17,6 | 4,0 | 5,3 | 5,9 | 6 | 8 | 9 |
| Исправление ошибок, доработка. С | 6 | 8 | 10 | 9 | 15 | 40 | 7,2 | 10,8 | 22 | 7,2 | 10,8 | 22,0 | 11 | 16 | 33 |
| Составление пояснительной записки Р,С | 14 | 6 | 6 | 20 | 12 | 12 | 16,4 | 8,4 | 8,4 | 8,2 | 4,2 | 4,2 | 12 | 6 | 6 |
| Итого: | 127 | 140 | 149 | 218 | 232 | 282 | 163,4 | 176,8 | 202,2 | 91,1 | 101,5 | 119,0 | 137 | 152 | 179 |

По данным расчетам, информационная технология для анализа данных будет разработана:

- В первом исполнении 137 дней
- Во втором исполнении 152 дней
- В третьем исполнении 179 дней

Следовательно, можно сделать вывод, что в первом исполнении работы будет выполнена быстрее.

На основе таблицы 4.6 был построен календарный план-график, представленный в таблице 4.7. График строился для максимального по длительности исполнения работ в рамках научно-исследовательского проекта с разбивкой по месяцам и декадам (10 дней) за период времени дипломирования. При этом работы на графике выделены различной штриховкой в зависимости от исполнителей, ответственных за ту или иную работу.

Таблица 4.7 - Календарный план-график проведения НИОКР по теме

| № | Вид работ | T _{кi} , кал. дн. | Продолжительность выполнения работ | | | | | | | | | | | | | | |
|-------|---|----------------------------------|------------------------------------|---|---|------|---|---|--------|---|---|-----|---|---|------|---|---|
| | | | февр. | | | март | | | апрель | | | май | | | июнь | | |
| | | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | |
| 1 | Продумать идею и записать основные данные. Р,С | 3 | ■ | | | | | | | | | | | | | | |
| 2 | Подробное описание идеи. Р,С | 5 | ■ | ■ | | | | | | | | | | | | | |
| 3 | Выявление функциональных блоков. Р,С | 2 | | ■ | | | | | | | | | | | | | |
| 4 | Определения содержания блоков. Р,С | 2 | | ■ | | | | | | | | | | | | | |
| 5 | Календарное планирование работ. Р | 5 | | ■ | | | | | | | | | | | | | |
| 6 | Поиск решения по каждому функциональному блоку. С | 5 | | | ■ | | | | | | | | | | | | |
| 7 | Реализация найденного решения. С | 44 | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| 8 | Тестирование. С | 11 | | | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| 9 | Отладка: устранение ошибок и повторное тестирование блоков проекта. С | 7 | | | | | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| 10 | Оценка эффективности полученных результатов. Р, С | 6 | | | | | | | | | | ■ | ■ | ■ | ■ | ■ | ■ |
| 11 | Определение целесообразности проведения ОКР. Р,С | 3 | | | | | | | | | | | ■ | ■ | ■ | ■ | ■ |
| 12 | Разработка диаграмм по проекту. С | 10 | | | | | | | | | | | ■ | ■ | ■ | ■ | ■ |
| 13 | Выбор проектируемой технологии. Р,С | 1 | | | | | | | | | | | | | | ■ | ■ |
| 14 | Оценка эффективности информационной технологии. Р,С | 3 | | | | | | | | | | | | | | ■ | ■ |
| 15 | Тестирование конечными пользователями. С | 6 | | | | | | | | | | | | | | ■ | ■ |
| 16 | Исправление ошибок, доработка интерфейса. С | 11 | | | | | | | | | | | | | | ■ | ■ |
| 17 | Составление пояснительной записки Р,С | 12 | | | | | | | | | | | | | | ■ | ■ |
| Итого | | | ■ – Студент ■ – Руководитель | | | | | | | | | | | | | | |

4.3. Бюджет научно-технического исследования (НТИ)

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используется следующая группировка затрат по статьям:

- Материальные затраты НТИ;
- Расчет амортизации;
- Основная заработная плата исполнителей темы;
- Дополнительная заработная плата исполнителей темы;
- Отчисления во внебюджетные фонды (страховые отчисления);
- Накладные расходы.

4.3.1. Расчет материальных затрат НТИ

Расчет материальных затрат осуществляется по формуле 4.7:

$$Z_m = (1 + k_T) \cdot \sum_{i=1}^m C_i \cdot N_{расч_i}, \quad (4.7)$$

где m – количество видов материальных ресурсов, потребляемых при выполнении научного исследования;

$N_{расч_i}$ – количество материальных ресурсов i -го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, м² и т.д.);

C_i – цена приобретения единицы i -го вида потребляемых материальных ресурсов (руб./шт., руб./кг, руб./м, руб./м² и т.д.);

k_T – коэффициент, учитывающий транспортно-заготовительные расходы. Транспортные расходы принимаются в пределах 15-25% от стоимости материалов

Так как работа предполагает непосредственное взаимодействие с персональным компьютером. Для разработки информационной технологии необходимы следующие материальные ресурсы:

- Системный блок;
- Принтер;

- Клавиатура;
- Монитор;
- Компьютерная мышь.

Расчет материальных затрат представлен в таблице 4.8.

Таблица 4.8 - Материальные затраты

| Наименование | Ед. изм. | Количество | | | Цена за ед., руб. | | | Затраты на материалы, (Зм), руб. | | |
|-------------------|----------|------------|-------|-------|-------------------|-------|-------|----------------------------------|--------------|--------------|
| | | Исп.1 | Исп.2 | Исп.3 | Исп.1 | Исп.2 | Исп.3 | Исп.1 | Исп.2 | Исп.3 |
| Системный блок | шт. | 1 | 1 | 1 | 38290 | 45800 | 47000 | 38290 | 45800 | 47000 |
| Принтер | | 1 | 1 | 1 | 6999 | 7999 | 5399 | 6999 | 7999 | 5399 |
| Монитор | | 1 | 2 | 2 | 9690 | 12200 | 13500 | 9690 | 24400 | 27000 |
| Клавиатура | | 1 | 1 | 1 | 1450 | 1520 | 1900 | 1450 | 1520 | 1900 |
| Компьютерная мышь | | 1 | 1 | 1 | 700 | 1200 | 890 | 700 | 1200 | 890 |
| Итого: | | | | | | | | 57129 | 80919 | 82189 |

4.3.2 Основная заработная плата исполнителей темы

Основная заработная плата рассчитывается по формуле 4.8:

$$Z_{осн} = Z_{дн} \cdot T_p, \quad (4.8)$$

где $Z_{осн}$ – основная заработная плата одного работника;

T_p – продолжительность работ, выполняемых научно-техническим работником, раб. дн. (табл. 8);

$Z_{дн}$ – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле 4.9:

$$Z_{дн} = \frac{Z_m \cdot M}{F_d}, \quad (4.9)$$

где Z_m – месячный должностной оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года. При отпуске в 28 раб. дня $M = 11,08$.

F_d – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн. (табл. 13).

Месячный должностной оклад работника рассчитывается по формуле 4.10:

$$Z_m = Z_{TC} \cdot (1 + k_p + k_d) \cdot k_r, \quad (4.10)$$

где Z_{TC} – заработная плата по тарифной ставке, руб.;

$k_{пр}$ – премиальный коэффициент, равный 0,3; (30% от Z_{TC})

k_d – коэффициент доплат и надбавок;

k_r – районный коэффициент, равный 1,3.

Результаты расчета действительного годового фонда проведены в таблице 4.9.

Таблица 4.9 – Баланс рабочего времени

| Показатели рабочего времени, дни | Исполнение 1 | Исполнение 2 | Исполнение 3 |
|--|--------------|--------------|--------------|
| Календарное число дней в году | 365 | 365 | 365 |
| Количество нерабочих дней | | | |
| Выходные | 104 | 104 | 104 |
| Праздники (фактически по каждому году) | 13 | 13 | 13 |
| Планируемые потери отпуска | 28 | 28 | 28 |
| Действительный годовой фонд | 220 | 220 | 220 |

Расчет затрат на основную заработную плату приведен в таблице 4.10.

При этом затраты на оплату труда студента-дипломника определяются как оклад студента 13952,44, а оклад руководителя проекта составляет 23264,86.

Коэффициент K , учитывающий районный коэффициент равный 1,3.

Таблица 4.10 - Затраты на основную заработную плату

| | Исполнители | Разряд | Z_{TC} , руб. | $k_{пр}$ | k_d | k_r | Z_m , руб. | $Z_{он}$, руб. | T_p , раб.дн | $Z_{осн}$, руб. |
|--------|--------------|-----------|-----------------|----------|-------|-------|--------------|-----------------|----------------|------------------|
| Исп. 1 | Руководитель | Ст. преп. | 23264,86 | 0,3 | 0,15 | 1,3 | 31605,3 | 1591,8 | 24,3 | 38679,7 |
| | Студент | 1 | 13952,44 | 0,3 | 0,15 | 1,3 | 18954,4 | 954,6 | 88,1 | 84101,3 |
| | Итого | | | | | | | | | 122781,1 |
| Исп. 2 | Руководитель | Ст. преп. | 23264,86 | 0,3 | 0,15 | 1,3 | 31605,3 | 1591,8 | 27,6 | 43932,5 |
| | Студент | 1 | 13952,44 | 0,3 | 0,15 | 1,3 | 18954,4 | 954,6 | 97,5 | 93074,7 |
| | Итого | | | | | | | | | 137007,2 |
| Исп. 3 | Руководитель | Ст. преп. | 23264,86 | 0,3 | 0,15 | 1,3 | 31605,3 | 1591,8 | 29,5 | 46956,9 |
| | Студент | 1 | 13952,44 | 0,3 | 0,15 | 1,3 | 18954,4 | 954,6 | 115 | 109780,4 |
| | Итого | | | | | | | | | 156737,3 |

4.3.3 Дополнительная заработная плата

Дополнительная заработная плата включает заработную плату за не отработанное рабочее время, но гарантированную действующим законодательством [40].

Расчет дополнительной заработной платы ведется по формуле 4.11:

$$Z_{доп} = k_{доп} \cdot Z_{осн}, \quad (4.11)$$

где $k_{доп}$ – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

$k_{доп}$ равен 0,12. Результаты по расчетам дополнительной заработной платы сведены в таблицу 4.11.

Таблица 4.11 - Затраты на дополнительную заработную плату

| Исполнитель | Основная зарплата(руб.) | | | Коэффициент дополнительной заработной платы ($k_{доп}$) | Дополнительная зарплата(руб.) | | |
|--------------|--------------------------|---------|----------|---|--------------------------------|-----------------|-----------------|
| | Исп.1 | Исп.2 | Исп.3 | | Исп.1 | Исп.2 | Исп.3 |
| Руководитель | 38679,7 | 43932,5 | 46956,9 | 0,12 | 4641,57 | 5271,90 | 5634,82 |
| Студент | 84101,3 | 93074,7 | 109780,4 | 0,12 | 10092,16 | 11168,96 | 13173,65 |
| Итого | | | | | 14733,73 | 16440,86 | 18808,47 |

4.3.4. Отчисления во внебюджетные фонды (страховые отчисления)

Величина отчислений во внебюджетные фонды определяется исходя из формулы 4.11:

$$Z_{внеб} = k_{внеб} \cdot (Z_{осн} + Z_{доп}), \quad (4.12)$$

где $k_{внеб}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На основании пункта 1 ст.58 закона №212-ФЗ для учреждений, осуществляющих образовательную и научную деятельность в 2014 году

вводится пониженная ставка страховых взносов –27,1%. Отчисления во внебюджетные фонды представлены в таблице 4.15.

Таблица 4.15- Отчисления во внебюджетные фонды

| Исполнитель | Основная заработная плата, руб. | | | Дополнительная заработная плата, руб. | | |
|--|---------------------------------|----------|-----------|---------------------------------------|----------|----------|
| | Исп. 1 | Исп. 2 | Исп. 3 | Исп. 1 | Исп. 2 | Исп. 3 |
| Руководитель проекта | 38679,73 | 43932,53 | 46956,87 | 4641,57 | 5271,90 | 5634,82 |
| Студент | 84101,32 | 93074,67 | 109780,38 | 10092,16 | 11168,96 | 13173,65 |
| Коэффициент отчислений во внебюджетные фонды | 27,1% | | | | | |
| Итого | | | | | | |
| Исполнение 1 | 37266,51 | | | | | |
| Исполнение 2 | 41584,42 | | | | | |
| Исполнение 3 | 47572,89 | | | | | |

4.3.5. Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Их величина определяется по формуле 4.12:

$$Z_{накл} = (\text{сумма статей} \div 7) \cdot k_{нр}, \quad (4.13)$$

где $k_{нр}$ – коэффициент, учитывающий накладные расходы.

Величина коэффициента накладных расходов равна 16%.

Исполнение 1: $(57129/7) \cdot 0,16 = 1305,81$

Исполнение 2: $(80919/7) \cdot 0,16 = 1849,6$

Исполнение 3: $(82189/7) \cdot 0,16 = 1878,61$

4.3.6. Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы (темы) является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции [41].

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведен в таблице 4.16.

Таблица 4.16 - Расчет бюджета затрат НТИ

| Наименование статьи | Сумма, руб. | | | Примечание |
|---|-------------|------------|------------|-------------|
| | Исп.1 | Исп.2 | Исп.3 | |
| 1. Материальные затраты НТИ | 57129 | 80919 | 82189 | Пункт 3.4.1 |
| 2. Затраты по основной заработной плате исполнителей темы | 122781,047 | 137007,203 | 156737,254 | Пункт 3.4.2 |
| 3. Затраты по дополнительной заработной плате исполнителей темы | 14733,73 | 16440,86 | 18808,47 | Пункт 3.4.3 |
| 4. Отчисления во внебюджетные фонды | 37266,51 | 41584,42 | 47572,89 | Пункт 3.4.4 |
| 5. Накладные расходы | 1305,81 | 1849,6 | 1878,61 | Пункт 3.4.5 |
| 6. Бюджет затрат НТИ | 233216,10 | 277801,08 | 307186,22 | Сумма 1-6 |

Вывод: Основываясь на данных, полученных в пунктах 4.3.1 – 4.3.5, был рассчитан бюджет затрат научно-исследовательской работы для трех исполнителей. Наиболее низким по себестоимости оказался проект первого исполнителя, затраты на его полную реализацию составляют 233216,10 рублей.

4.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Эффективность определяется на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с

определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трех (или более) вариантов исполнения научного исследования.

Интегральный финансовый показатель разработки определяется по формуле 4.13:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}, \quad (4.14)$$

где $I_{\text{финр}}^{\text{исп.}i}$ – интегральный финансовый показатель разработки;

Φ_{pi} – стоимость i -го варианта исполнения;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

Расчет:

$$I_{\text{финр}}^{\text{исп}1} = \frac{233216,10}{307186,22} = 0,75,$$

$$I_{\text{финр}}^{\text{исп}2} = \frac{277801,08}{307186,22} = 0,90,$$

$$I_{\text{финр}}^{\text{исп}3} = \frac{307186,22}{307186,22} = 1,$$

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное увеличение бюджета затрат разработки в разгах (значение больше единицы), либо соответствующее численное удешевление стоимости разработки в разгах (значение меньше единицы, но больше нуля).

Как видно из результатов, полученных выше, Исполнение 1 имеет наименьший интегральный финансовый показатель, а это значит, что данное исследование обладает наименьшей стоимостью разработки.

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить по формуле 4.14:

$$I_{pi} = \sum a_i \cdot b_i, \quad (4.15)$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i^a, b_i^p – бальная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности приведен в таблице 4.17.

Таблица 4.17 - Сравнительная оценка характеристик вариантов исполнения проекта

| Объект исследования Критерии | Весовой коэффициент параметра | Исп.1 | Исп.2 | Исп.3 |
|---|-------------------------------|-------------|-------------|------------|
| 1. Способствует росту производительности труда пользователя | 0,3 | 5 | 3 | 4 |
| 2. Удобство в эксплуатации (соответствует требованиям потребителей) | 0,15 | 4 | 2 | 3 |
| 3. Помехоустойчивость | 0,15 | 5 | 3 | 3 |
| 5. Надежность | 0,25 | 4 | 4 | 4 |
| 6. Материалоемкость | 0,15 | 4 | 4 | 4 |
| ИТОГО: | 1 | 4,45 | 3,25 | 3,7 |

$$I_{p-исп1} = 5*0,3 + 4*0,15 + 5*0,15 + 4*0,25 + 4*0,15 = 4,45;$$

$$I_{p-исп2} = 3*0,3 + 2*0,15 + 3*0,15 + 4*0,25 + 4*0,15 = 3,25;$$

$$I_{p-исп3} = 4*0,3 + 3*0,15 + 3*0,15 + 4*0,25 + 4*0,15 = 3,7.$$

Исходя из вычислений, показатель ресурсоэффективности Исполнения 1 имеет достаточно высокое значение (по 5-бальной шкале), что говорит об эффективности использования технического проекта. Высокие баллы надежности и помехоустойчивости позволяют сделать вывод о надежности разрабатываемой информационной технологии. НТИ-2 и НТИ-3 уступают НТИ-1 в данном показателе, что свидетельствует о низкой ресурсоэффективности данных исследований в сравнении с Исполнением 1.

Интегральный показатель эффективности **вариантов исполнения разработки** ($I_{исп.i}$) рассчитывается по формуле 4.15

$$I_{исп.1} = \frac{I_{p-исп1}}{I_{финр.1}}, \quad (4.16)$$

$$I_{исп.1} = \frac{I_{p-исп1}}{I_{финр.1}} = \frac{4,45}{0,75} = 5,93;$$

$$I_{исп.2} = \frac{I_{p-исп1}}{I_{финр.1}} = \frac{3,25}{0,90} = 3,6;$$

$$I_{исп.3} = \frac{I_{p-исп1}}{I_{финр.1}} = \frac{3,7}{1} = 3,7.$$

Опираясь на полученные данные, можно сделать вывод, что эффективность исполнения у НТИ-1 выше, чем у других (равно 5,93), на втором месте НТИ-3 (3,7), и самый низкий показатель у НТИ-2 (3,6).

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволяет определить сравнительную эффективность проекта и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта (\mathcal{E}_{cp}) рассчитывается по формуле 4.16:

$$\mathcal{E}_{cp} = \frac{I_{исп.1}}{I_{исп.2}}, \quad (4.17)$$

$$\mathcal{E}_{cp1} = \frac{5,93}{3,6} = 1,64, \quad \mathcal{E}_{cp2} = \frac{3,6}{3,7} = 0,97, \quad \mathcal{E}_{cp3} = \frac{3,7}{5,93} = 0,62,$$

Сравнительная эффективность разработки, представлена в таблице 4.18.

Таблица 4.18 - Сравнительная эффективность разработки

| № п/п | Показатели | Исп.1 | Исп.2 | Исп.3 |
|-------|---|-------|-------|-------|
| 1 | Интегральный финансовый показатель разработки | 0,75 | 0,90 | 1 |
| 2 | Интегральный показатель ресурсоэффективности разработки | 4,45 | 3,25 | 3,7 |
| 3 | Интегральный показатель эффективности | 5,93 | 3,6 | 3,7 |
| 4 | Сравнительная эффективность вариантов исполнения | 1,64 | 0,97 | 0,62 |

Общий вывод по разделу:

В ходе работы был определен целевой рынок проекта, построена карта сегментирования, выявлено какие ниши на рынке услуг не заняты конкурентами или где уровень конкуренции низок.

Произведена оценка качества и перспективности по технологии QuaD, где было выявлено, что данная разработка является перспективной.

Произведен SWOT-анализ, в ходе которого были рассмотрены слабые и сильные стороны разработки информационной технологии, выявлены возможные перспективы ее создания и рассмотрены варианты минимизации влияния угроз, которые могут этому помешать. Для повышения эффективности работы необходимо минимизировать ошибки, контролировать процесс создания и проверять работу информационной технологии в соответствии с поставленными требованиями.

В рамках анализа конкурентных технических решений, определена трудоемкость данного проекта, она равна 137 дням. Это на 15 дней меньше, чем у НТИ-2, и на 42 дня меньше, чем у НТИ-3, что говорит о том, что данный проект будет выполнен быстрее, чем его конкурентные варианты. На основании трудоемкости была построена диаграмма Ганта.

Так же были рассчитаны величины затрат научно-исследовательских работ. В результате проведенных расчетов, бюджет затрат НТИ Исполнения 1 составил 233216,10. Бюджет НТИ Исполнения 2 равен 277801,08. 3 Исполнения: 307186,22. Из чего можно сделать вывод о том, что первый исполнитель наименее затратный при реализации проекта.

Первое исполнение лидирует и по показателю сравнительной эффективности, оно равно 1,64, в то время как для НТИ-2 этот показатель равен 0,97, а для НТИ-3 - 0,62. С позиций технической и финансовой ресурсоэффективности можно сделать выводы о том, что научно – техническое решение, представленное первым исполнителем, является более предпочтительным.

Глава 5 Социальная ответственность

Направление данной ВРК – разработка информационной технологии для анализа медицинских данных. Работа заключается во взаимодействии научных работников с компьютером. Рабочей зоной является кабинет №204 инженерной школы информационных технологий и робототехники Национального Исследовательского Томского Политехнического Университета. Рабочим местом научных работников являются оборудованный компьютерный стол и офисный стул.

Данный раздел предназначен для разработки комплекса мер технического, организационного, режимного и правового характера, минимизирующих негативные последствия разработки информационной технологии.

5.1 Анализ выявленных вредных факторов

Выявлены следующие негативные факторы:

- Недостаточная освещенность рабочей зоны;
- Электромагнитные излучения
- Микроклимат.

Недостаточная освещенность рабочей зоны

Правильно спроектированное и выполненное производственное освещение способствует повышению производительности труда, снижает утомляемость, улучшает условия зрительной работы, благотворно влияет на производственную среду, оказывая положительное психологическое воздействие на работающего, снижает травматизм и повышает безопасность труда.

Недостаточность освещения приводит к напряжению зрения, ослабляет внимание, приводит к наступлению преждевременной утомленности. Чрезмерно яркое освещение вызывает ослепление, раздражение и резь в глазах. Неправильное направление света на рабочем месте может создавать

резкие тени, блики, дезориентировать работающего. Превышение нормативных параметров освещения ведет к снижению работоспособности, так как чрезмерная яркость и блескость слепит глаза и искажает видимость.

Помещение, освещенное посредством естественного и искусственного света одновременно, называется смешанным. Нормы искусственного и естественного освещения указаны в документе СНиП 23-05-95 [42]. Освещенность на поверхности стола в зоне размещения рабочего документа должна быть равна 300 - 500 лк. Освещение не должно создавать бликов на поверхности экрана. Все описанные показатели должны соответствовать нормам СанПиН 2.2.2/2.4.1340-03[43].

Для обеспечения требуемой освещенности рассчитаем систему освещения на рабочем месте.

Для рассматриваемого помещения наиболее разумно выбрать систему общего равномерного освещения, данная система используется для тех помещений, где работа производится на всей площади и нет необходимости в лучшем освещении отдельных участков.

В качестве источников света рационально использовать люминесцентные лампы, т. к. они имеют много преимуществ: их спектр близок к естественному свету, они имеют большую экономичность, светоотдачу и срок службы. Однако наряду с этим имеются и недостатки: их работа иногда сопровождается шумом. Они хуже работают при низких температурах; их нельзя применять во взрывоопасных помещениях; имеют малую инерционность. Тип светильников для люминесцентных ламп – трехламповый светильник типа ЛПО, т. к. они предназначены для освещения в нормальных помещениях, а параметры микроклимата помещения, в котором осуществляется работа по ГОСТ 30494-96 «Параметры микроклимата в помещениях» соответствуют категории нормального помещения".

Значения нормируемой освещенности изложены в строительных нормах и правилах СНиП 23-05-95. Нормами для данных работ установлена необходимая освещённость рабочего места $E=300$ лк, соответствующая

зрительной работе очень высокой точности (наименьший размер объекта различения равен 0.15 – 0.3 мм разряд зрительной работы – II, подразряд зрительной работы – Г, фон – светлый, контраст объекта с фоном – большой). Полученная величина освещенности корректируется с учетом коэффициента запаса по причине загрязнения светильников и уменьшения светового потока ламп.

Расчёт системы освещения производится методом коэффициента использования светового потока, который выражается отношением светового потока, падающего на расчётную поверхность, к суммарному потоку всех ламп.

Основные характеристики используемого осветительного оборудования и рабочего помещения:

Наименьшая допустимая высота подвеса для светильников типа ЛПО с люминесцентными лампами выбираем из СНиП 23-05-95, и она равна – 2,5 метра, основные работы производятся на высоте 0,75 метра над поверхностью пола. Таким образом, $h = 2,5 - 0,75 = 1,75$ м – высота подвеса светильника над рабочей поверхностью. Выбираем значение λ из СНиП 23-05-95, $\lambda = 1,1$. Следовательно, расстояние между светильниками $L = 1,1 \times 1,75 = 1,9$ м. Расстояние от стен помещения до крайних светильников - $L/3=0,7$ м. Размер светильников тип ЛПО: $A=0,595$ м, $B=0,595$ м.

Количество рядов светильников с люминесцентными лампами определяется по формуле (5.1):

$$n_{\text{ряд}} = \frac{(B - \frac{2}{3}L)}{L} + 1 = \frac{(0,595 - \frac{2}{3} \cdot 1,9)}{1,9} + 1 = 2,5 \approx 3, \quad (5.1)$$

Где $n_{\text{ряд}}$ – количество рядов;

B – ширина помещения, м;

L – расстояние между рядами светильников, м.

Количество светильников с люминесцентными лампами определяется по формуле (5.2):

$$n_{ce} = \frac{(A - \frac{2}{3}L)}{l_{ce} + 0,5} = \frac{(6 - \frac{2}{3} \cdot 1,9)}{0,595 + 0,5} = \frac{4,67}{1,095} = 4,3 \approx 5, \quad (5.2)$$

Где n_{ce} – количество светильников в ряду;

A – длина помещения, м;

l_{ce} – длина светильника, м.

Общее количество светильников с люминесцентными лампами в помещении определяется по формуле (5.3):

$$N = n_{ряд} \cdot n_{ce} = 15, \quad (5.3)$$

Найдем индекс помещения по формуле (5.1):

$$i = \frac{S}{h \cdot (A + B)} = \frac{36}{1,75 \cdot (12)} = \frac{36}{21} = 1,71, \quad (5.4)$$

где S – площадь помещения, м²;

h – высота подвеса светильников над рабочей поверхностью, м;

A, B – длина и ширина помещения.

Величина светового потока лампы определяется по следующей формуле (6.5):

$$\Phi = \frac{E \cdot k \cdot S \cdot Z}{n \cdot \eta} = \frac{300 \cdot 1,5 \cdot 36 \cdot 1,1}{45 \cdot 0,38} = \frac{17820}{17,1} = 1042 \text{ лм}, \quad (5.5)$$

Где Φ – световой поток каждой из ламп, Лм;

E – минимальная освещенность, Лк, k – коэффициент запаса (1.5), S – площадь помещения, м² (36 м²), n – число ламп в помещении (45), η – коэффициент использования светового потока (Тогда для светильников типа ЛПО $\eta=0,38$), Z – коэффициент неравномерности освещения (для светильников с люминесцентными лампами $Z=0,9$).

Световой поток равен 1303 лм. Из СНиП 23-05-95 выбираем ближайшую по мощности стандартную лампу. В нашем случае это лампа ЛД мощностью 30 Вт.

Определим электрическую мощность осветительной установки:

$$P = 45 \cdot 30 = 1350 \text{ Вт}$$

Таким образом, оптимальной для рассматриваемого помещения будет

система освещения, состоящая из 15 светильников типа ЛПО, каждый из которых имеет 3 люминесцентные лампы ЛД мощностью 30 Вт. Мощность осветительной установки составит 1350Вт.

Электромагнитные излучения

Электромагнитные поля, представляют реальную угрозу. Влияние электромагнитного излучения вызывает изменение обмена веществ на клеточном уровне, нарушение деятельности сердечно-сосудистой и центральной нервной системы, нарушаются биологические процессы в тканях и клетках, также воздействует на органы зрения. Беременным женщинам рекомендуется избегать большого влияния электромагнитных излучений. Электромагнитные излучения повреждают защитное биополе человека и отрицательно сказываются на его здоровье, повышая риск возникновения различного рода заболеваний [44]. Напряженность электрического поля и плотность магнитного потока представлены в таблицах 6.1, 6.2 соответственно

Таблица 6.1 – Напряженность электрического поля

| Диапазон частот | Напряженность электрического поля |
|---------------------|-----------------------------------|
| 5 Гц — 2 кГц, E1 | 25 В/м |
| 2 кГц — 400 кГц, E2 | 2,5 В/м |

Таблица 6.2 – Плотность магнитного потока

| Диапазон частот | Плотность магнитного потока |
|---------------------|-----------------------------|
| 5 Гц — 2 кГц, B1 | 250 нТл |
| 2 кГц — 400 кГц, B2 | 25 нТл |

Для того, чтобы свести электромагнитное излучение к минимуму, системные блоки необходимо располагать настолько дальше от сотрудника, насколько это возможно, так же применяются жидкокристаллические мониторы, излучение которых намного ниже. Мониторы должны быть расположены так, чтобы излучения поглощались не людьми, то есть в углу или у стены.

Предельно допустим уровнем напряженности электрической и магнитной составляющих является диапазон частот 30 кГц - 300 МГц. Уровень

напряженности электрической и магнитной составляющих на рассматриваемом месте находятся на допустимом уровне.

Микроклимат.

Непостоянная температура в помещении может вызвать физический дискомфорт у сотрудников, что негативно повлияет на их производительности труда. При высокой температуре воздуха в помещении кровеносные сосуды поверхности тела расширяются. При понижении температуры окружающего воздуха кровеносные сосуды кожи сужаются.

Повышенная влажность ($\varphi > 85\%$) затрудняет терморегуляцию вследствие снижения испарения пота, а очень низкая влажность ($\varphi < 20\%$) вызывает пересыхание слизистых оболочек дыхательных путей.

Санитарные правила и нормы 2.2.4.548-96 устанавливают определённый микроклимат для всех типов рабочих помещений. Они выполняют обязательную функцию для любой из возможных форм собственности. СанПиН 2.2.4.548-96 обязывает следовать безусловным правилам и требованиям, выполнение которых согласно статье 212 ТК РФ напрямую зависит от работодателя [45]. Невыполнение требований ведёт к ответственности и закрепляется как нарушение трудового законодательства и ущемление прав, работающих [46].

Категории работ и нормы указаны ниже в таблице 6.3. Допустимые нормы представлены в таблице 6.4.

Таким образом, для теплового самочувствия человека важно определенное сочетание температуры, относительной влажности и скорости движения воздуха. Данные были взяты из СанПиНа 2.2.4.548-96.

Таблица 6.3 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений

| Период года | Категория работ по уровню энергозатрат, Вт | Температура воздуха, °С | Температура поверхностей, °С | Относительная влажность воздуха, % | Скорость движения воздуха, м/с |
|-------------|--|-------------------------|------------------------------|------------------------------------|--------------------------------|
| Холодный | Ia (до 139) | 22 - 24 | 21 - 25 | 60 - 40 | 0,1 |
| Теплый | Ia (до 139) | 23 - 25 | 22 - 26 | 60 - 40 | 0,1 |

Таблица 6.4 – Допустимые величины показателей микроклимата

| Период года | Категория работ по уровню энергозатрат, Вт | Температура воздуха, °С | | Температура поверхности, С | Относительная влажность воздуха, % | Скорость движения воздуха, м/с | |
|-------------|--|-----------------------------------|-----------------------------------|----------------------------|------------------------------------|---|---|
| | | Диапазон ниже оптимальных величин | Диапазон выше оптимальных величин | | | Для диапазона температур воздуха ниже оптимальных величин, не более | Для диапазона температур воздуха выше оптимальных величин, не более** |
| Холодный | Ia (до 139) | 20,0-21,9 | 24,1-25,0 | 19,0-26,0 | 15-75* | 0,1 | 0,1 |
| Теплый | Ia (до 139) | 21,0-22,9 | 25,1-28,0 | 20,0-29,0 | 15-75* | 0,1 | 0,2 |

Температура воздуха в помещении в диапазоне от 23 до 26 °С. Влажность воздуха – около 40 %. Скорость движения воздуха – 0,1 м/сек. Достигается это за счет использования средств местного кондиционирования воздуха, кроме этого помещение проветривается во время обеденного перерыва. Охлаждение сотрудников не происходит, так как помещение хорошо отапливается.

Параметры микроклимата рассматриваемого помещения соответствуют нормам, следовательно, соблюдаются необходимые для работы условия.

Шум

Шум в окружающей среде создается одиночными или комплексными источниками, находящимися снаружи или внутри здания. Шум – это совокупность звуков различной интенсивности и частоты (шелест, дребезжание, скрип, визг и т. п.) [47]. Шум может создаваться оборудованием, людьми, городским транспортом. Источниками шумовых помех могут быть: компьютер и периферийное оборудование к нему, вентиляционные установки и кондиционеры. Предельно допустимый уровень звука представлен в таблице 6.5.

Таблица 6.5 – Предельно допустимый уровень звука по СН 2.2.4/2.1.8.562-96

| Рабочие места | Уровень звука, дБА |
|---|--------------------|
| Конструкторские бюро, программисты, лаборатории | 50 |
| Помещения управления, рабочие комнаты | 60 |

Основным источником шума в кабинете являются вентиляторы блоков питания. Уровень шума колеблется от 35 до 40дБА.

Шумовая обстановка на рабочем месте соответствует норме. Регулярные проверки оборудования позволят исключить превышения допустимого уровня шума.

5.2 Анализ психофизических факторов

Психофизические факторы также оказывают большое влияние на продуктивность выполняемых работ работником. В результате анализа выделены следующие психофизические факторы:

- Монотонность труда;
- Статические нагрузки;
- Перенапряжение зрения;
- Умственное перенапряжение.

Монотонность труда;

Монотонную называют такую работу, отличительными признаками которой являются однообразие рабочих действий, их многократное повторение и небольшая длительность. Если операция сводится к выполнению ограниченного круга трудовых действий, то она является монотонной даже при значительной длительности.

Монотонная работа оказывает существенное влияние на функциональное состояние человека. Под влиянием монотонности человек становится вялым, безучастным к работе, возникает угнетающее состояние, что приводит к преждевременному утомлению. Утомление сказывается в том, что человек чувствует снижением своей работоспособности, даже когда производительность труда еще не падает. Это снижение работоспособности выражается в переживании особого, тягостного напряжения и в неуверенности человек чувствует, что не в силах должным образом продолжить работу.

В конечном итоге она отрицательно сказывается и на результативные показатели трудовой деятельности: снижается производительность труда и

качественные его параметры. Монотонная работа снижает творческую активность работников и их интерес к выполняемой работе [48].

Статические нагрузки;

Человек, работающий за персональным компьютером, обычно сидит в неподвижной позе и напряжен. Это приводит к усталости и появлению болей в шее, плечевых суставах, позвоночнике. Длительная работа за клавиатурой может вызывать неприятные болевые ощущения в запястьях, локтевых суставах, кистях и пальцах рук.

Перенапряжение зрения;

Работа с монитором – является самым вредным фактором, воздействующим на пользователя. Длительное нахождение перед дисплеем компьютера вредно отражается на зрении человека – повышается раздражимость, снижает его остроту, ухудшается работа головного мозга – у человека могут возникать головные боли и резь в глазах. После четырёх часов работы за компьютером начинаются недомогания. Наименьшую нагрузку на глаза оказывает считывание информации с монитора, чуть больше – ввод информации, а самая сильная утомляемость возникает при работе в диалоговом режиме и занятия (просмотр) компьютерной графикой. Причинами нарушения зрения являются мерцание экрана (возникает при частоте горизонтальной развертки не выше 50-60 Гц), невысокая резкость символов, мелкий текст программ и документов (нормируется отношение ширины знака к высоте) [49], наличие бликов и искажений, проблемы с оптимальным соотношением яркости и контрастности, неправильно выбранная освещенность в помещении. Все это создает серьезные проблемы, приводящие к зрительному дискомфорту и ухудшению зрения у 60-85 % пользователей.

Для предупреждения ухудшения зрения необходимо руководствоваться следующим:

Использовать дисплей с достаточно большим размером экрана и частотой горизонтальной развертки свыше 70 Гц;

Устранить блики с экрана:

- Монитор должен иметь матовую одноцветную поверхность;
- Не располагать экран, обращенным в сторону окна (допускается размещать монитор перпендикулярно окну);
- Пользоваться оконными шторами, занавесками, жалюзи;
- Рабочее место не располагать непосредственно под источником верхнего искусственного освещения;
- Стена напротив экрана монитора должна быть матовой;

Умственное перенапряжение.

Умственный труд – это работы, связанные с приёмом и переработкой информации, требующие напряжения сенсорного аппарата внимания, памяти, активации процессов мышления [50].

Умственный труд имеет ряд особенностей. Чаще всего он связан с сидячим образом жизни и длительной работой в закрытом помещении. Усиленная работа мозга требует большого притока крови к нему, что в свою очередь связано с повышением тонуса сосудов мозга. Это физиологическое повышение тонуса сосудов при неправильной организации труда может перейти в патологическое, что может привести к стойкому повышению артериального давления.

Как всякая интенсивная работа, умственный труд неизбежно связан с утомлением, которое приводит к тому, что человеку приходится затрачивать на обычный объем работы больше энергии. Несмотря на это, он допускает ошибки, и, наконец, утомление приводит к невозможности продолжать работу. При неумении правильно организовать умственный труд наступает состояние хронического утомления, которое может закончиться истощением нервной системы или развитием сосудистого заболевания.

Для того что бы избежать негативного воздействия психофизических факторов необходимо сменить ритм трудовой деятельности, устраивать 5-10 минутные перерывы в работе приблизительно через каждый час, а во время перерыва рекомендуется использовать активный отдых. Это позволит снизить

утомляемость и поднять работоспособность человека. Учитывая отрицательное влияние монотонной работы на работоспособность работников очень важно использовать с учетом конкретных складывающихся условий все возможные меры предупреждения этого нежелательного явления. Особое значение здесь имеет правильное проектирование содержания трудовых операций, установление и соблюдение рационального режима труда и отдыха.

Для наиболее интенсивной и трудной умственной работы лучше отводить утренние часы. Приступая к работе, необходимо позаботиться о том, чтобы комната, в которой предстоит работать, в достаточной степени вентилировалась. Источник освещения рабочего места должен находиться слева и не быть слишком ярким, чтобы не вызвать утомления глаз. Шум мешает работе, поэтому его по возможности нужно исключить.

5.3 Анализ выявленных опасных факторов

Так как работа сотрудника связана непосредственно с компьютером, следовательно, подвержена воздействию опасных факторов производственной среды. Этими факторами являются:

- Механические опасности;
- Термические опасности;
- Электробезопасность;

Рассмотрим более подробно каждый из указанных выше факторов.

Механические опасности, возникающие при работе с ПК в офисе, могут вызвать тяжелые предметы быта, такие как:

- Стулья;
- Столы;
- Системные блоки;
- Мониторы;
- Мультимедийное оборудование больших габаритов;
- Вентиляторы;
- Радиаторы отопления.

При работе с описанными выше предметами, необходимо соблюдать простые инструкции:

- Не пытаться починить оборудование без соответствующих знаний;
- Не переставлять мебель и оборудование без согласования с руководством;
- Отключать оборудование перед перемещением на другую позицию.
- Термические опасности в офисе могут быть связаны со следующими причинами:

- Радиаторы отопления;
- Источники бесперебойного питания ПК;
- Компьютеры.

Для того, чтобы бороться с термическими опасностями нужно соблюдать следующие правила:

- Не сливать воду с радиаторов отопления;
- Не разбирать любую технику во включенном состоянии;
- Осторожно пользоваться приборами быта при разогреве пищи и приготовления напитков с горячей водой.

Электробезопасность является опасным фактором и обычно она связана со следующими источниками:

- Статическое электричество;

Статическое электричество – это совокупность явлений, связанных с возникновением, сохранением свободного электрического заряда на поверхности материалов изделий или на изолированных проводниках. Заряды накапливаются на оборудовании и материалах, а сопровождающие электрические разряды могут явиться причиной пожаров и взрывов, нарушения технологических процессов, точности показаний электрических приборов и средств автоматизации.

Допустимые уровни напряженности электростатических полей установлены в ГОСТ 12.1.045-84 «Электростатические поля. Допустимые уровни на рабочих местах и требования к проведению контроля.» Допустимые

уровни напряженности полей зависят от времени пребывания на рабочих местах. Предельно допустимый уровень напряженности электростатических полей равен 60 кВ/м в 1 ч.

Применение средств защиты работающих обязательно в тех случаях, когда фактические уровни напряженности электростатических полей на рабочих местах превышают 60 кВ/м [51].

- Электрический ток;

Электрический ток представляет собой скрытый тип опасности, т.к. его трудно определить в токо- и нетокопроводящих частях оборудования, которые являются хорошими проводниками электричества. Смертельно опасным для жизни человека считают ток, величина которого превышает 0,05А, ток менее 0,05А – безопасен (до 1000 В).

В рассматриваемом рабочем месте, находятся применяемые в работе компьютеры, принтер, которые представляют собой опасность повреждения переменным током. Источники постоянного тока на рабочем месте отсутствует.

Общие травмы, вызванные действием электрического тока – электрический удар, могут привести к судорогам, остановке дыхания и сердечной деятельности. Местные травмы: металлизация кожи, механические повреждения, ожоги, также очень опасны. (ГОСТ12.1.006-84 «Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля»)

Мероприятия защиты при электробезопасности следующие:

- Отключать электрооборудование при его ремонте;
- Периодически снимать электростатическое напряжение, касаясь пальцами рук, заземленных поверхностей;
- Для безопасности во время гроз необходимо удостовериться о наличие молниеотвода, и того факта что все розетки в офисном помещении заземлены.

5.4 Экологическая безопасность

Рассматривается рабочее место на исследуемом предприятии, которое занимается образовательной деятельностью. Характер производственной деятельности не предполагает наличие стационарных источников загрязнения окружающей среды. Источников загрязнения атмосферы нет.

При рассмотрении влияния персонального компьютера на атмосферу и гидросферу вредных выбросов не выявлено. Анализ воздействия на литосферу сводится к обычному бытовому мусору и отбросам жизнедеятельности человека. В случае выхода из строя компьютеров, они списываются и отправляются на специальный склад, который при необходимости принимает определенные меры по утилизации списанной техники и комплектующих.

5.5 Безопасность в чрезвычайных ситуациях

На таком объекте как офисное помещение могут возникнуть такие чрезвычайные ситуации (ЧС) как:

- Техногенные;
- Экологические;
- Природные.

Пожары.

Пожары – это наиболее часто встречающаяся чрезвычайная ситуация в офисных помещениях. Под пожаром понимается неконтролируемое горение во времени и пространстве, наносящее материальный ущерб и создающее угрозу жизни и здоровью людей. Огнетушительные вещества: вода, песок, пена, газообразные вещества, порошок, инертные газы. Основными причинами такой ЧС как пожар являются повреждения или замыкания электропроводки, оборудование, обрыв проводов, не соблюдению мер пожаробезопасности в офисе [52].

Для того, чтобы избежать возникновения пожара необходимо проводить следующие профилактические работы, направленные на устранение возможных источников возникновения пожара:

- Проведение инструктажа офисных работников о пожаробезопасности;
- Периодическая проверка проводки.

Общие требования к пожарной безопасности нормируются ГОСТ 12.1.004–91.

Пожаробезопасность.

Рассматриваемое помещение оснащено средствами пожаротушения в соответствии с нормами. На 100 м² пола имеется:

- Пенный огнетушитель ОП-10 – 1 шт.;
- Углекислотный огнетушитель ОУ-5 – 1 шт.;
- Ящик с песком на 0,5 м³ – 1 шт.;
- Железные лопаты – 2 шт.

Помещение, в котором проводится работа, оборудовано противопожарной сигнализацией и датчиками дыма. Компьютерное оборудование для работы исправно. Пожарные гидранты, пожарный водопровод и средства пожаротушения исправны и находятся на своих штатных местах в состоянии готовности к работе.

В соответствии с общероссийскими нормами технологического проектирования все производственные здания и помещения по взрывопожарной опасности подразделяются на категории А, Б, В, Г и Д. Рассматриваемый кабинет по взрывопожароопасной подходит под категорию В. Согласно СП 12.13130.2009 в зимнее время гидранты утеплены, пожарный водопровод заизолирован и утеплен, и не разморожен [53].

В рассматриваемом кабинете приказом назначается лицо, отвечающее за соблюдение правил пожарной безопасности, за исправное состояние пожарного инвентаря и за применение первичных способов пожаротушения.

Краны противопожарного водопровода оснащены брезентовыми шлангами с брандспойтами. Соединительные головки кранов и шлангов должны иметь резиновые прокладки. Скрученные прорезиненные шланги и брандспойты хранятся в опломбированных шкафчиках, размещенных вблизи кранов.

Ящики и щиты, где хранится противопожарный инвентарь, ручки лопат и пожарных топоров, окрашены в красный цвет, а металлические части периодически смазываются и очищаются для предотвращения коррозии.

В случае возникновения ЧС как пожар, необходимо предпринять меры по эвакуации персонала из офисного помещения в соответствии с принятым планом эвакуации (рисунок 5.1). При невозможности самостоятельно потушить пожар необходимо вызвать пожарную команду, после чего поставить в известность о случившемся инженера по технике безопасности.

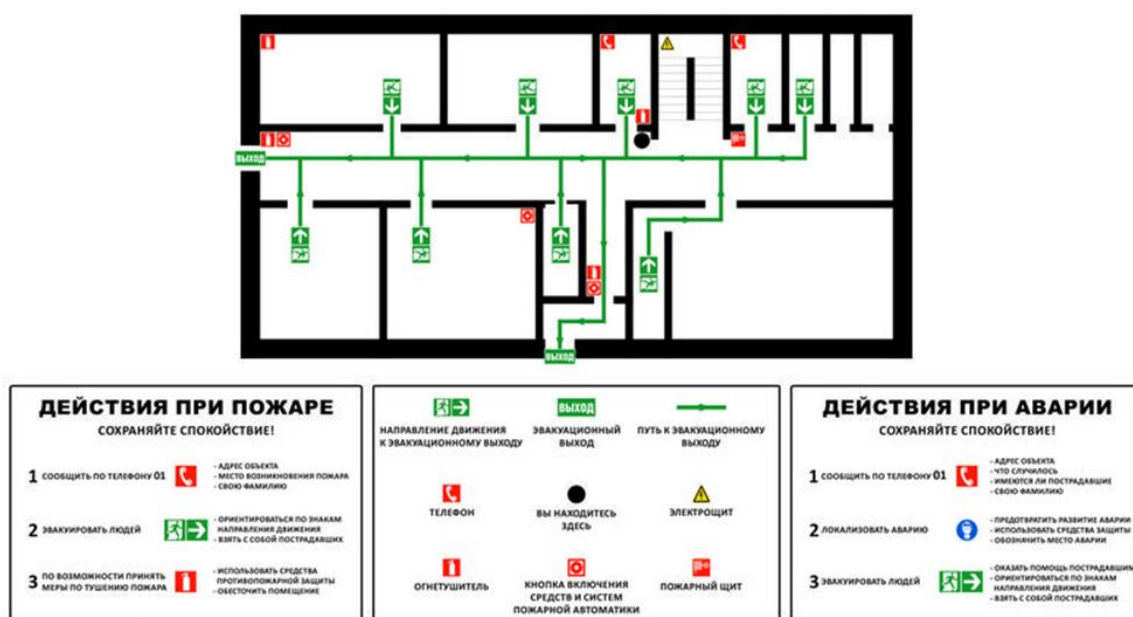


Рисунок 5.1 - План эвакуации

Землетрясения.

Согласно единой схеме распределения землетрясений на земном шаре, Западная Сибирь входит в число сейсмически спокойных материковых областей, то есть где почти никогда не бывает землетрясений с магнитудой разрушительной величины свыше 5 баллов.

Здание инженерной школы информационных технологий и

робототехники Национального Исследовательского Томского Политехнического Университета относится к кладке С (обычное качество, устойчивость к горизонтальной нагрузке проектом здания не предусмотрена). Таким образом, можно сделать вывод, что землетрясения не угрожают.

5.6 Правовые и организационные вопросы обеспечения безопасности

Регулирование отношений в сфере охраны труда в Томской области осуществляется Трудовым кодексом Российской Федерации, федеральными законами и иными нормативными правовыми актами Российской Федерации, содержащими нормы трудового права. Государственное управление охраной труда на территории Томской области осуществляют федеральные органы исполнительной власти, Администрация Томской области.

В соответствии с государственными стандартами и правовыми нормами обеспечения безопасности предусмотрена рациональная организация труда в течение смены, которая предусматривает:

- Длительность рабочей смены не более 8 часов;
- Установление двух регламентируемых перерывов (не менее 20 минут после 1-2 часов работы, не менее 30 минут после 2 часов работы);
- Обеденный перерыв не менее 40 минут. Обязательно предусмотрен предварительный медосмотр при приеме на работу и периодические медосмотры. Каждый сотрудник должен пройти инструктаж по технике безопасности перед приемом на работу и в дальнейшем, должен быть пройден инструктаж по электробезопасности и охране труда.

Требования к организации рабочих мест пользователей:

- Рабочее место должно быть организовано с учетом эргономических требований в соответствии с ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» и ГОСТ 12.2.061-81 «ССБТ. Оборудование производственное. Общие требования безопасности к рабочим местам»;

- Конструкция рабочей мебели (рабочий стол, кресло, подставка для ног) должна обеспечивать возможность индивидуальной регулировки соответственно росту пользователя и создавать удобную позу для работы. Вокруг ПК должно быть обеспечено свободное пространство не менее 60 - 120см

Вывод: в ходе выполнения данного раздела были рассмотрены и описаны аспекты производственной безопасности, аспекты экологической безопасности, меры безопасности в чрезвычайных ситуациях, правовые и организационные вопросы обеспечения безопасности. Изложены требования к рабочему месту сотрудника. Созданные условия должны обеспечивать комфортную работу. На основании изученной литературы, были указаны оптимальные параметры микроклимата, а также проведен выбор системы и расчет оптимального освещения производственного помещения. Соблюдение условий, определяющих оптимальную организацию рабочего места научного сотрудника, позволит сохранить хорошую работоспособность в течение всего рабочего дня, что в свою очередь будет способствовать быстрой разработке информационной технологии.

Заключение

В результате выполнения выпускной квалификационной работы, была разработана информационная технология анализа результатов лечения детей с различными формами ожирения, на основе анализа клинико-лабораторных показателей.

С использованием разработанной информационной технологии было осуществлено восстановление пропусков в исходных наборах данных и проанализированы результаты лечения пациентов с использованием визуальных образов основанных на таком типе пиктограмм, как лица Чернова.

В рамках выполнения ВКР была построена структурная схема алгоритма информационной технологии и осуществлена её практическая реализация.

В ходе исследования была выявлена результативность лечения пациентов пяти различных групп исходя из которой лечение в группе два и пять не оказало должного воздействия.

Апробация информационной технологии проводилась под контролем и в сотрудничестве с НИИ курортологии и физиотерапии г. Томска, все выводы согласованы со специалистами.

В разделе «Социальная ответственность» рассмотрены основные параметры по безопасности рабочего места инженерной школы информационных технологий и робототехники ТПУ, которые полностью соответствует нормам СанПиН.

Разработанная информационная технология имеет широкую сферу применения и может использоваться для комплексного анализа данных в различных медицинских учреждениях.

Результаты работы представлены в следующих научных публикациях и выступлениях на конференциях:

1. Публикация в журнале «Информационные технологии в науке, управлении, социальной сфере и медицине»: И.А. Лызин Использование

визуализации при анализе медицинских данных // Информационные технологии в науке, управлении, социальной сфере и медицине. Сборник научных трудов IV международной научной конференции в 2 ч., Томск, 5-8 декабря 2017. - Томск: ТПУ С.339-342.

2. Участие в VIII международном молодёжном научном форуме «Новые форматы транснациональной научно-образовательной деятельности» с докладом «Информационная технология анализа результатов лечения детей с различными формами ожирения» 26 – 27 апрель 2018. ТГПУ диплом III степени.

3. IV Международная научно-практическая конференция молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук» (статья «Информационная технология анализа результатов лечения детей с различными формами ожирения» включена в работу конференции и принята в печать).

4. III Всероссийская конференция с международным участием «Здоровье и качество жизни» (статья «Разработка алгоритма анализа результатов лечения детей с эндокринопатиями» включена в работу конференции и принята в печать).

Список используемых источников

1. Всемирная организация здравоохранения / Ожирение и избыточный вес [электронный ресурс] – режим доступа: <http://www.who.int/mediacentre/factsheets/fs311/ru/>.
2. Самойленко Н. Э., Кувина В. Н., Кувин С. С. Комплексный анализ медицинских данных // Вестник ВГТУ. 2009. №9. С.114-118
3. Кучерявский С. В., Беляев И. А. Применение методов анализа многомерных данных для классификации клеток крови // Известия АлтГУ. 2007. №1. С. 73-75.
4. Струков Денис Раймондович, Горохов Владимир Леонидович Геоинформационные системы и многомерные статистические методы пространственного анализа для исследования заболеваемости // Информационно-управляющие системы. 2009. №3. С. 52-54
5. Осадчая И.А., Берестнева О.Г. Методы исследования структуры многомерных экспериментальных данных // Научное обозрение. Технические науки. – 2014. – № 2. – С. 92-92;
6. Лосев Александр Георгиевич, Зенович Андрей Васильевич, Бочкарев Олег Андреевич, Левшинский Владислав Викторович Интеллектуальный анализ многомерных термометрических данных в медицинской диагностике // Вестник ВолГУ. Серия 1: Математика. Физика. 2016. №5 (36). С. 150-161
7. Осадчая Ирина Анатольевна, Берестнева Ольга Григорьевна, Немеров Евгений Владимирович Анализ многомерных медицинских данных с помощью пиктографиков «Лица Чернова» // Бюллетень сибирской медицины. 2014. №4. С. 89-93.
8. Стромов Г.Г., Рыжков Д.В. Исследование границ применимости методики интегральной оценки в задачах анализа многомерных медицинских данных// XX Международная научно-практическая конференция «СОВРЕМЕННЫЕ ТЕХНИКА И ТЕХНОЛОГИИ» С. 357-358

9. Романова, И.К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах [электронный ресурс] – режим доступа: <http://elibrary.ru/item.asp?id=25837825>

10. Шаропин Константин Александрович, Берестнева Ольга Григорьевна, Воловоденко Виталий Алексеевич, Марухина Ольга Владимировна Визуализация медицинских данных на базе пакета NovoSpark // Известия ЮФУ. Технические науки. 2010. №8. С. 242-248.

11. Герасимова А.С. Визуализация многомерных данных с помощью естественной поверхности // Приволжский научный вестник. 2013. №2 (18). С. 15-17.

12. В.К.Финн, В.Г.Блинова, Е.С.Панкратова, Е.Ф.Фабрикантова ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ ДЛЯ АНАЛИЗА МЕДИЦИНСКИХ ДАННЫХ. Часть 1 // Врач и информационные технологии. 2006. №5. С. 62-70

13. Беребин Михаил Алексеевич Методология и практика разработки методик клинической (медицинской) психодиагностики на основе применения экспертного метода, методик обработки экспертных оценок и методов многомерного анализа данных // Вестник ЮУрГУ. Серия: Психология. 2010. №27 (203). С. 9-12.

14. Кокоулин А.Н., Южанинов Р.И. Многомерный анализ данных по обращаемости в лечебные учреждения с помощью средств Oracle olap // Вестник ПНИПУ. Электротехника, информационные технологии, системы управления. 2015. №13. С 5-14.

15. Никонорова М.Л. Интеллектуальный анализ медицинских данных с использованием кейсовой технологии // Врач и информационные технологии. 2016. №1. С 54-59.

16. А.И. Минжасова Статистический анализ медицинских данных [электронный ресурс] – режим доступа: <http://elibrary.ru/item.asp?id=24843191>

17. В.В. Быкова, А.В. Катаева Методы и средства анализа информативности признаков при обработке медицинских данных // Программные продукты и системы. 2016. №2 (114). С. 172-178.
18. Кожевников С. Н., Денисов А. В., Новикова И. И., Ерофеев Ю. В. Использование многомерных методов статистического анализа для определения программ лечебно-профилактической помощи медицинским работникам в современных условиях // ВНМТ. 2013. С 147-148.
19. Куликова К. Ю. Статистический анализ медицинских данных при наличии пропусков [электронный ресурс] – режим доступа: <https://elibrary.ru/item.asp?id=23096578>.
20. О. В. Марухина, О. Г. Берестнева, Е. Е. Мокина Визуализация и анализ многомерных экспериментальных данных [электронный ресурс] – режим доступа: <http://elibrary.ru/item.asp?id=27432149>.
21. Захаров А.А., Оленников Е.А., Паюсова Т.И., Петелина Т.И., Мусихина Н.А., Гапон Л.И., Осипова И.В., Такканд А.Г., Белослудцева О.Е. Научный анализ данных в медицинской информационной системе (на примере определения факторов, влияющих на уровень с-реактивного белка, с помощью нейронных сетей) // Вестник Тюменского государственного университета. Социально-экономические и правовые исследования. 2014. №7. С 251-527
22. Осадчая И. А., Прокопьев Р. О. Технологии визуализации в задачах оценки функционального состояния // ИВД. 2014. №4-2.
23. Jolita Bernatavičienė, Gintautas Dzemyda, Olga Kurasova, Virginijus Marcinkevičius, Viktor Medvedev The Problem of Visual Analysis of Multidimensional Medical Data [электронный ресурс] – режим доступа: https://link.springer.com/chapter/10.1007/978-0-387-36721-7_17
24. Science Direct/
R.A.RobbD.P.HansonR.A.KarwoskiA.G.LarsonE.L.WorkmanM.C.Stacy
[электронный ресурс] – режим доступа:
<https://www.sciencedirect.com/science/article/pii/0895611189902851>

25. Science Direct / Noor Elaiza AbdKhalidMarinaYusoffEzzatul AkmaKamaru-ZamanIzyan IzzatiKamsani Multidimensional Data Medical Dataset Using Interactive Visualization Star Coordinate Technique [электронный ресурс] – режим доступа: <https://www.sciencedirect.com/science/article/pii/S1877050914014975>
26. Anbarasi.M.S Outlier Detection for Multidimensional Medical Data [электронный ресурс] – режим доступа: <http://ijcsit.com/docs/Volume%202/vol2issue1/ijcsit2011020109.pdf>
27. Волкова В.Н., Васильев А.Ю., Ефремов А.А., Юрьев В.Н. Классификация информационных технологий // Открытое образование. 2015. №5
28. Всемирная организация здравоохранения / Ожирение и избыточный вес [электронный ресурс] – режим доступа: <http://www.who.int/mediacentre/factsheets/fs311/ru/>.
29. К.А. Шаропин, О.Г. Берестнева, Г.И. Шкатова ВИЗУАЛИЗАЦИЯ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ / Известия Томского политехнического университета. Инжиниринг георесурсов. – 2010 – С. 172.
30. Паклин Н. Б., Орешков В. И. Визуализация данных. Бизнес-аналитика. От данных к знаниям. Питер, 2013.
31. Яшков Ю.И. Этапы развития хирургии ожирения //Вестник хирургии. – 2003. –№ 3.
32. Берестнева О.Г. Методы структурного анализа и визуализации экспериментальных данных в социальных и медицинских исследованиях / О.Г. Берестнева, И.А. Осадчая, А.Л. Бурцева. - Томск: Из-во Томского политехн. ун-та, 2014 – С. 17-19.
33. Роберт И. Кабаков R в действии. Анализ и визуализация данных в R / Метод множественного восстановления данных / Москва – 2014 – С. 443
34. Гублер Е. В. Вычислительные методы анализа и распознавания патологических процессов – МЕДИЦИНА, 1978. – 198с.

35. The R Project for Statistical Computing [электронный ресурс] – режим доступа: <https://www.r-project.org/>
36. MATLAB [электронный ресурс] – режим доступа: <https://matlab.ru/products/matlab>.
37. IBM / IBM SPSS Software [электронный ресурс] – режим доступа: <https://www.ibm.com/analytics/ru/ru/technology/spss/>
38. SAS / Advanced Analytics [электронный ресурс] – режим доступа: https://www.sas.com/ru_ru/solutions/analytics.html
39. RapidMiner Studio [электронный ресурс] – режим доступа: <https://rapidminer.com/>.
40. И.Г. Видяев, Г.Н. Серикова, Н.А. Гаврикова // Финансовый менеджмент, ресурсоэффективность и ресурсосбережение: учебно-методическое пособие - Издательство Томского политехнического университета 2014.
41. Методы менеджмента качества: Методы поиска новых идей и решений / Е.А. Кузьмина, А.М. Кузьмин. - Т., 2003. - 187 с.
42. Естественное и искусственное освещение [Электронный ресурс] / Система нормативных документов в строительстве; ред. ЖурбаВ. С.; Web-разработчик Зайцева Е. П. - Электрон, дан. - М.: Освещение., 2007. URL: http://www.tehbez.ru/Docum/DocumShow_DocumID_312.html/ , - Загл. с экрана.— Яз.рус. Дата обращения: 28.05.2015 г.
43. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.
44. Электромагнитные поля на рабочем месте [Электронный ресурс] / Пособие по безопасности жизнедеятельности; ред. Мухина Е. С.; Web-разработчик Плашкевич Е. О. - Электрон, дан. - М.: Электромагнит., 2004. URL: <http://habrahabr.ru/post/140431/>, -Загл. с экрана.— Яз.рус. Дата обращения: 27.05.2015 г.
45. Производственная и экологическая безопасность: Методические указания по разработке раздела выпускной квалификационной работы для

студентов всех форм обучения/[Электронный ресурс] Сост. М.Э. Гусельников, В.Н. Извеков, Н. В. Крепша, В.Ф. Панин. – Томск.,2010. – 265 с.

46. Микроклимат[Электронный ресурс] / Журнал «Охрана труда и техника безопасности» №5; ред. Корниенко К. И.; Разработчик Гнедов Н. А. - Электрон, дан. - М.: БЖД., 2011. URL: <http://www.otb.by/articles/mikroklimat/> - Загл. с экрана. – Яз.рус. Дата обращения: 26.05.2015 г.

47. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки [Электронный ресурс] / Журнал «Охрана труда и техника безопасности» №6; ред. Корниенко К. И.; Разработчик Гнедов Н. А. - Электрон, дан. - М.: БЖД., 2011. URL: <http://www.vashdom.ru/sanpin/224-218562-96/> - Загл. с экрана. – Яз.рус. Дата обращения: 27.05.2015 г.

48. Монотонность труда [Электронный ресурс] – режим доступа: <http://kursak.net/monotonnost-truda/>.

49. . Санитарно-эпидемиологические правила и нормативы СанПиН 2.2.2/2.4. 1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организация работ.

50. Умственный труд, его физиологические особенности. Меры профилактики умственного утомления [Электронный ресурс] – режим доступа: <http://fbuz19.ru/about/news/detail.php?ID=736>

51. Бутиков Е.И., Быков А.А., Кондратьев А.С. Физика для поступающих в вузы. - М., 1991. - с.328.

52. Правила пожарной безопасности в Российской Федерации. – М.: Министерство Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий, 2003.

53. СП 12.13130.2009 Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности (с Изменением N 1) Госкомсанэпиднадзор, 2009.

**Приложение А
(справочное)**

Разделы:
Введение
Обзор литературы
Глава 1. Постановка задачи
Глава 2. Выбор решения задачи

Студент:

| Группа | ФИО | Подпись | Дата |
|--------|--------------------------|---------|------|
| 8КМ61 | Лызин Иван Александрович | | |

Консультант ОИТ:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|------------|--------------------------------|---------------------------|---------|------|
| Доцент ОИТ | Чердынцев Евгений Сергеевич | к.т.н. | | |

Консультант – лингвист ОИЯ:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|-----------------------------|---------------------------|---------------------------|---------|------|
| Ст. преподаватель ОИЯ | Краснова Татьяна Ивановна | | | |

Introduction

The problem of overweight is called one of the illnesses of civilization as its prevalence in the modern society constantly increases. Recently, statistical observations have recorded an increase in the frequency of obesity in children and teenagers. According to the World Health Organization, over the past four decades in the world, the proportion of children and teenagers (aged 6 to 17) having obesity has increased tenfold.

Currently, there are many approaches to the treatment of obesity, accumulated arrays of clinical-laboratory indicators of patients. In this connection the task of developing information technologies for the rapid processing of these arrays and obtaining new knowledge becomes urgent. Under information technology is meant a process that uses a combination of means and methods of collecting, processing and transmitting data to obtain new quality information about the state of an object, process or phenomenon.

The purpose of the work is to create the information technology for assessment of treatment results for five different groups of patients, suffering from obesity of various degrees. Approaches to treatment in each group are changing.

The final qualification work is devoted to ways, helping to find the solution to the following tasks:

- Analysis of the literature and internet sources to get acquainted with the subject domain and experience of solving similar tasks;
- Choice of methods and tools for solving the task;
- Assessment of the initial data set of clinical-laboratory indicators – identifying missing data and research of regularities of emergence of missing values;
- Preparation of data for analysis – imputation missing data and creation of the full data set;
- Creation of visual images before and after treatment on the basis of recovered data;

- Coordination of research results with specialists from the Scientific Research Institute of Balneology and Physiotherapy, formation of outputs;
- Calculation of resource efficiency and resource saving;
- Development of a set of technical regulations and legal measures, minimizing the negative consequences of the development of information technology.

Literature review

Table 1. Analysis of publications on the research theme

| № | Authors, article title | Organization | Annotation |
|---|---|--|--|
| 1 | Jolita Bernatavičienė, Gintautas Dzemyda, Olga Kurasova, Virginijus Marcinkevičius, Viktor Medvedev The Problem of Visual Analysis of Multidimensional Medical Data [1]. | Biodynamics Research Unit, Mayo Foundation, Rochester | The problem is considered of visual analysis of the multidimensional medical data. Also the article describes combining such methods of data analysis as classification and visualization. This synthesis lets us obtain a more objective conclusions on the analyzed data. |
| 2 | R.A. Robb, D.P. Hanson, R.A. Karwowski, A.G. Larson, E.L. Workman, M.C. Stacy Analyze: A Comprehensive, operator-interactive software package for multidimensional medical image display and analysis [2]. | Institute of Mathematics and Informatics Vilnius, Lithuania | The work contains a description of a comprehensive software package, called ANALYZE, which has been developed and permits detailed investigation and evaluation of multidimensional biomedical images. ANALYZE can be used with 3-D imaging modalities based on x-ray computed tomography, radionuclide emission tomography, ultrasound tomography, and magnetic resonance imaging. The software is written entirely in "C" and runs on standard UNIX workstations. |
| 3 | Noor Elaiza, Abd Khalid, MarinaYusoff, Ezzatul Akma, Kamaru-Zaman Izyan Izzati, Kamsani Multidimensional Data Medical Dataset Using Interactive Visualization Star Coordinate Technique [3]. | University Technology MARA, Shah Alam, Malaysia | This article explores the possibility of an interactive method of visualizing coordinates to determine the correlation of clusters between the selected attributes using interactive stars coordinates for multidimensional data sets. |
| 4 | Anbarasi M.S. Outlier Detection for Multidimensional Medical Data [4]. | Pondicherry Engineering College | The main theme of the article is devoted to the problem of outliers detection in medical data. Outlier detection is an important research problem that aims to find objects that are considerably dissimilar, exceptional and inconsistent in the database. |

Chapter 1. Problem definition

1.1 Basic concepts

Data analysis is the process of checking, cleaning, transforming and modeling data in order to identify useful information. Data analysis has many aspects and approaches, covering the different methods under different names, in different areas of business and social sciences.

Multidimensional data is data that contains information about two or more characteristics for each studied object.

Multidimensional data is characterized by:

- High dimension of attribute space caused by several dozens of characteristics;
- Presentable statistics missing;
- Missed data;
- Interval nature of data caused by condition uncertainty of getting it.

Analysis of multidimensional data is an approach based on the application of mathematical methods in which phenomena, processes and systems, are characterized by a set of variables.

The phrase «Medical data» is used in both a narrow and broad sense:

The Medical data in a broad sense are any data related to medicine in general.

In a narrow sense these are data related directly to a person as a patient, that is, information about his or her health, the characteristics of the body. We will adhere further to such understanding of medical data.

1.2 Description of subject domain

Obesity in children and teenagers is a topical and very serious problem. The prevalence, which is increasing: 25% of teenagers in the world are overweight, and 15% are obese. Overweight in childhood is a significant predictor of obesity in adulthood. Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health [5].

There are two main factors that can lead to childhood obesity: alimentary and endocrine.

The alimentary factor is associated with malnutrition or a sedentary lifestyle. The child overeats, but does not consume the stored energy.

The endocrine factor is associated first of all with the violation of the work of the endocrine system which is responsible for the work of internals. Obesity can lead to the following failures:

- Genetic predisposition.
- Hormonal violations. This is a violation work of the endocrine glands

Overweight and obesity are major risk factors for a number of chronic diseases such as diabetes, cardiovascular diseases, and cancer.

We obtained a dataset contained the values of clinical laboratory indicator groups of patients in the context of cooperation with Research Institute of Balneology and Physiotherapy in Tomsk. The condition of each patient is assessed when admitting to a treatment and at the end of treatment. The patients are children and teenagers at the age of 6 to 17 with overweight problems. The database contains data of 276 patients divided into 5 groups and by 106 indicators. The groups associate patients that receive medical treatment by different methods.

The current problem is solved on the basis of the following indicators:

WM – waist measurement of the patient

HW – hip width

Weight – The weight of the patient, kg

BMI – Body mass index

SBP mm Hg – Systolic blood pressure

DBP mm Hg. – Diastolic blood pressure

HOMA – Homeostasis model assessment of insulin resistance

Exercise tolerance

General lipid (4-8 g/l) – General blood lipids

Triacyglyceride (less than 1.7 mmol/L) – Triacyglyceride level in blood serum

Cholesterol (less than 4 mmol/L) – General cholesterol level in blood serum

Alpha cholesterol (more than 1,03 mmol/L) – Alpha cholesterol level in blood serum

VLDL (mmol/L) – Very low density lipoprotein

LDL (mmol/L) – Low-density lipoprotein

Atherogenicity index

Alkaline phosphatase – Alkaline phosphatase in blood serum

Catalase – Catalase of blood serum

MDA – Malondialdehyde in blood serum

Sialic acids – Sialic acids level in blood serum

Ceruloplasmin – Ceruloplasmin level in blood serum

Glucose (less than 5,6 mmol/L) – Blood glucose level on an empty stomach

T3 (1,0-2,8) – Triiodothyronine level in blood serum

FT3 (4-8.6) – Free triiodothyronine level

T4 (53-158) – Tetraiodothyronine level

FT4 (10-23.2) – Free tetraiodothyronine level

TSH (0,23-3,4) – Thyroid stimulating hormone level in blood serum

Anti-TPO – Thyroperoxidase antibodies in blood serum

Insulin – Insulin level in blood serum

Cortisol – Cortisol level in blood serum

Leptin – Leptin level in blood serum

Serotonin – Serotonin level in blood serum

IL-6 – Interleukine-6 level in blood serum

TNF (not > 2.5 pg / ml) – Tumor necrosis factor level in blood serum

IgA (0,7-3,0 g/l) – Concentration of immunoglobulin A level in blood serum

IgG (8,0-16,0 g/l) – Concentration of immunoglobulin G level in blood serum

IgM (0,8-2,5 g/l) – Concentration of immunoglobulin M level in blood serum

CIC (40-100) – Circulating immune complexes in blood serum

Lysozyme (40-60 y.e.) – Activity of lysozyme in blood serum

Kallikrein – Kallikrein level in blood serum

Prothrombin index – Prothrombin index in blood serum

Macroglobulin – Macroglobulin in blood serum

ACE – Angiotensin-converting enzyme in blood serum

NO – Nitric oxide level in blood

ox LDL – Oxidation low-density lipoprotein in blood serum

Chapter 2 Choice of the task completion

2.1 Choice of technology and the description of methods of the task completion

2.1.1 Data mining Technology

The term «Data Mining» is often translated as the data search, excavation or «extraction of knowledge grains from the data mountains». The Data mining technology is defined rather precisely by one of founders of this direction Grigory Piatetsky-Shapiro. «Data Mining» is a process of detection in raw data of knowledge earlier unknown, uncommon, practically useful and available to interpretation that are necessary for decision-making in various spheres of human activity.

Data Mining is a set of various methods, knowledge detection. The choice of method often depends on the type of available data and on the kind of information needs to be obtained.

The Data Mining methods and algorithms are following: artificial neural networks, decision trees, symbolical rules, the nearest neighbor and the k-nearest neighbor methods, the method of support vectors, Bayesian networks, linear regression, correlation and regression analysis; hierarchical methods of the cluster analysis, not hierarchical methods of the cluster analysis, including algorithms of k-means and k-median; methods of search of associative rules, including algorithm a priori; method of limited search, evolutionary programming and genetic algorithms, various methods of data visualization and set of other methods.

After studying methods and algorithms a method of multiple restoration of missed data will be used for resolving this problem of missed data restoring. Method of principal components, component (or factor) analysis and method of meaningless characteristics elimination will be considered as the methods of dimension lowering. Chernoff faces were chosen as a mean of visualization.

2.1.2 Analysis of outliers

Outliers are data points lying far away from the majority of other data points. Outliers of the data that is not normally distributed do not require identification. As most statistical tests assume that data are normally distributed, outlier identification should precede data analysis. Usually to identify the outliers used box plots. In this box plot, any data that lies outside the upper or lower fence lines is considered outliers. This graph simplifies the data, leaving only a few parameters: the median, upper and lower quartiles, significant minimum and maximum (smallest and largest values, not counting outliers) and also outliers which are substantially smaller or larger than the remaining numbers in the series.

There are basically three methods for treating outliers in a data set.

- The first method is to remove outliers as a means of trimming the data set. Under this approach, a data set that excludes outliers is analyzed.
- The second method involves replacing the values of outliers or reducing the influence of outliers through outlier weight adjustments. This approach involves modifying the weights of outliers or replacing the values being tested for outliers with expected values.
- The third method is used to estimate the values of outliers using robust techniques. When the nature of the population distributions is known, this approach is considered appropriate because it produces estimators robust to outliers, and estimators are consistent [4,6].

2.1.3 Analysis of missing data

Missing values and outliers are frequently encountered during the data collection phase of observational or experimental studies conducted in all fields of natural and social sciences.

Missing values can arise from information loss as well as dropouts and non-responses of the study participants. The presence of missing values leads to a smaller sample size than intended and eventually compromises the reliability of the study

results. It can also produce biased results when inferences about a population are drawn based on such a sample, undermining the reliability of the data. As a part of the pretreatment process, missing data are either ignored in favor of simplicity or replaced with substituted values estimated with a statistical method. In general, the analysis of missing values involves the consideration of efficiency, handling of missing data and the resulting complexity in analysis, and the bias between missing and observed values. There are three types of missing data:

- Missing completely at random (MCAR);
- Missing at random (MAR);
- Not missing at random (NMAR).

Too many missed values are the problem for further processing and analysis of data. Usually, 15% from the entire data extraction is the maximum percentage value of missed data. If the amount of percentage value exceeds 15%, you should discard this indicator [7].

Methods for handling missing values:

- Complete case analysis

This method uses only the data of variables observed at each time point for analysis after removing all missing values. While the simplicity of analysis is an advantage, reduced sample size and lower statistical power are disadvantages because drawing statistical inferences becomes difficult during analysis. It is the most commonly used method in statistical analysis programs such as SPSS and SAS to handle missing values.

- Available case analysis

This technique deals with only the data available for each analysis. It allows a larger sample size than that used for complete case analysis. However, this approach causes sample sizes to vary between the variables used in analysis.

- Imputation analysis

Imputation involves replacing missing values with substituted values obtained from a statistical analysis to produce a complete data set without missing values for analysis. Imputations can be created by using either an explicit or an implicit

modeling approach. The explicit modeling approach assumes that variables have a certain predictive distribution and estimates the parameters of each distribution, which is used for imputations. It includes different methods of imputation by mean, median, probability, ratio, regression, predictive-regression, and assumption of distribution. The implicit modeling approach focuses on computing an algorithm required to generate accurate imputation values, if possible.

2.1.2 Method of multiple imputation missing data

The method of multiple imputation of missed data is the way of filling missed data by means of iterative simulation. Multiple imputation is frequently used for dealing with missed data in complex cases. The Monte Carlo methods are used for filling missed values in derived data sets.

Standard statistical methods are applied to each of the derived data sets. Based on their results the assessment of results and confidential interval that take into account in determinedness caused by missed data are formed [8].

2.1.3 Pictographs «Chernoff Faces»

The Chernoff faces, invented by Herman Chernoff in 1973, display multidimensional data in the shape of a human face. The individual parts, such as eyes, ears, mouth and nose represent values of the variables by their shape, size, placement and orientation. The idea behind using faces is that people can easily recognize faces and notice small changes without difficulty.

2.1.4 The methods for lowering the dimensional of multidimensional data

Principal component analysis is the way of data dimension lowering that converts a large number of correlated variables into a much smaller one of uncorrelated variables called principal components.

The first main component (2.1)

$$PC_1 = a_1X_1 + a_2X_2 + \dots + a_kX_k, \quad (2.1)$$

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. The model looks so (2.2):

$$X_i = a_1F_1 + a_2F_2 + \dots + a_pF_p + U_i, \quad (2.2)$$

Where X_i – the i -th observable variable ($i = 1 \dots k$),

F_j – common factors ($j = 1 \dots p$), and $p < k$.

U_i – a unique constituent variable.

The method of Kullback

The method of Kullback is offers as a measure of informativeness measure the differences between the two classes, called divergence.

According to this method Kullback divergence, or information content is calculated by the formula (2.3):

$$I(x_j) = \sum_{i=1}^G [P_{i1} - P_{i2}] \cdot \log_2 \frac{P_{i1}}{P_{i2}}, \quad (2.3)$$

G – number of G -gradation characteristic;

P_{i1} – the probability of the i -th grades in first grade.

$$P_{i1} = \frac{m_{i1}}{\sum_{i=1}^G m_{i1}}, \quad (2.4)$$

m_{i1} – frequency of occurrence of the i -th grades in the first class;

The denominator is the appearance of all grades in first grade, that is, the total number of observations in the first grade.

P_{i2} – the probability of the i -th grades in the second grade (2.5).

$$P_{i2} = \frac{m_{i2}}{\sum_{i=1}^G m_{i2}} \quad (2.5)$$

m_{i2} – frequency of occurrence of the i -th grades in the second grade.

2.2. Choice of the decision tool

2.2.1. Scripting language R

R is the universal programming language developed for application in such areas as prospecting data analysis, classical statistical tests and high-level graphics. The language was created as similar to the language S developed in Bell Labs and was his alternative realization though there are essential differences between the languages, but generally the code works at language S in the environment R. Initially R has been developed by Ross Ihaka (English Ross Ihaka) and Robert Gentleman (English Robert Gentleman) (the first letter of their names – R); which are the staff of statistical faculty of University of Auckland. R Foundation organization supports and develops the language and the environment.

Thanks to the extensive and continuously extending library of packages, language R takes the leading positions in statistics, data analysis and data production.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS [9].