

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа информационных технологий и робототехники  
 Направление подготовки 09.04.04 Программная инженерия  
 Отделение школы (НОЦ) Отделение информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
<b>Алгоритмическое и программное обеспечение анализа и визуализации результатов медицинских осмотров</b>

УДК 004.415.2:004.421:616-071.2

Студент

Группа	ФИО	Подпись	Дата
8ПМ8И	Кащеева Евгения Викторовна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Аксёнов С.В.	к.т.н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Меньшикова Е.В.	к.ф.н., доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Горбенко М.В.	к.т.н.		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Губин Е.И.	к.ф.-м.н.		

## Планируемые результаты обучения по направлению

### 09.04.04 «Программная инженерия»

Код результата	Результат обучения (выпускник должен быть готов)
<b>Общие по направлению подготовки 09.04.04 «Программная инженерия»</b>	
P1	Проводить научные исследования, связанные с объектами профессиональной деятельности
P2	Разрабатывать новые и улучшать существующие методы и алгоритмы обработки данных в информационно-вычислительных системах
P3	Составлять отчеты о проведенной научно-исследовательской работе и публиковать научные результаты
P4	Проектировать системы с параллельной обработкой данных и высокопроизводительные системы
P5	Осуществлять программную реализацию информационно-вычислительных систем, в том числе распределенных
P6	Осуществлять программную реализацию систем с параллельной обработкой данных и высокопроизводительных систем
P7	Организовывать промышленное тестирование создаваемого программного обеспечения
<b>Профиль «Технологии больших данных» / «Big data solutions»</b>	
P8	Исследовать и анализировать большие данные, создавать их модели и интерпретировать структуры данных в таких моделях
P9	Понимать принципы создания, хранения, управления, передачи и анализа больших данных с использованием новейших технологий, инструментов и систем обработки данных в высокопроизводительных сетях
P10	Применять теорию распределенной системы управления базами данных к традиционным распределенным системам реляционных баз данных, облачным базам данных, крупномасштабным системам машинного обучения и хранилищам данных



	медицинским университетом.
<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><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>Аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; анализ и визуализация результатов медицинских осмотров; построение классификатора степеней тяжести заболевания пациентов с рожистым воспалением; определение наиболее важных для классификации признаков; обсуждение результатов выполненной работы; заключение работы.</p> <p>Дополнительно должны быть разработаны следующие разделы: финансовый менеджмент, ресурсоэффективность и ресурсосбережение; социальная ответственность; раздел на иностранном языке.</p>
<p><b>Перечень графического материала</b></p> <p><i>(с точным указанием обязательных чертежей)</i></p>	
<p><b>Консультанты по разделам выпускной квалификационной работы</b></p> <p><i>(с указанием разделов)</i></p>	
<b>Раздел</b>	<b>Консультант</b>
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Меньшикова Екатерина Валентиновна
Социальная ответственность	Горбенко Михаил Владимирович
Раздел на иностранном языке	Пичугова Инна Леонидовна
<p><b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b></p>	
Литературный обзор	

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	
---	--

**Задание выдал руководитель / консультант (при наличии):**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Аксёнов С.В.	к.т.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ПМ8И	Кащеева Евгения Викторовна		

## РЕФЕРАТ

Выпускная квалификационная работа 90 с., 10 рис., 26 табл., 26 источников, 3 прил.

Ключевые слова: медицинские документы, текстовые данные, классификация, определение важности признаков, градиентный бустинг

Объектом исследования является (ются) медицинская документация, формируемая на основе осмотров пациентов лечащим врачом.

Цель работы – разработка алгоритмического и программного обеспечения анализа и визуализации результатов медицинских осмотров

В процессе исследования проводились аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; анализ и визуализация результатов медицинских осмотров; построение классификатора степеней тяжести заболевания пациентов с рожистым воспалением; определение наиболее важных для классификации признаков.

В результате исследования проведены анализ и визуализация результатов медицинских осмотров; разработана модель классификации степеней тяжести заболевания пациентов с рожистым воспалением; определены наиболее важные для классификации признаки.

Область применения: здравоохранение

Экономическая эффективность/значимость работы применение системы позволит автоматически определять степени тяжести заболевания, что сократит время работы специалиста, затрачиваемое на данную процедуру.

В будущем планируется работа по внедрению алгоритма в медицинские системы заполнения документации

## Оглавление

Введение.....	9
1 Обзор литературы .....	11
2 Объект и методы исследования .....	14
2.1 Объект исследования .....	14
2.2 Методы исследования.....	16
2.2.1 Метод определения нормальной формы слова.....	16
2.2.2 Метод определения веса признаков документа «TF-IDF».....	16
2.2.3 Метод дерева решений .....	18
2.2.4 Метод градиентного бустинга .....	19
3 Расчеты и аналитика .....	21
3.1 Выбор программного обеспечения .....	21
3.2 Загрузка и предварительный анализ данных .....	21
3.3 Предварительная подготовка, построение «TF-IDF» матрицы и определение значений меток классов .....	24
3.4 Разделение набора данных на тестовую и тренировочную выборки .....	25
3.5 Построение классификатора .....	25
3.5.1 Выбор модели классификации.....	25
3.5.2 Подбор оптимальных параметров классификатора.....	26
3.5.3 Построение классификатора с оптимальными параметрами .....	27
4 Результаты.....	28
4.1 Классификатор степеней тяжести рожистого воспаления.....	28
4.2 Определение важности признаков .....	29
5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	33
5.1 Предпроектный анализ .....	34
5.1.1 Технология «QuaD» .....	34
5.1.2 Диаграмма Исикавы.....	35
5.1.3 Оценка готовности научно-исследовательского проекта к коммерциализации .....	36
5.2 Инициация научно-исследовательского проекта.....	38

5.2.1 Цели и результаты научно-исследовательского проекта.....	38
5.2.2 Организационная структура научно-исследовательского проекта.....	39
5.3 Планирование управления научно-исследовательским проектом.....	39
5.3.1 План научно-исследовательского проекта .....	39
5.3.2 Бюджет научно-исследовательского проекта .....	40
5.3.2.1 Расчет материальных затрат .....	41
5.3.2.2 Расчет затрат на электроэнергию, мобильную связь и интернет.....	41
5.3.2.3 Амортизационные затраты.....	42
5.3.2.4 Заработная плата исполнителей .....	42
5.3.2.5 Отчисления во внебюджетные фонды (страховые отчисления).....	44
5.3.2.6 Накладные расходы.....	44
5.3.2.7. Формирование бюджета затрат научно-исследовательского проекта ...	44
5.3.3 Риски научно-исследовательского проекта.....	45
5.3.4 Описание потенциального эффекта .....	45
6 Социальная ответственность .....	49
Введение.....	49
6.1 Правовые и организационные вопросы обеспечения безопасности .....	49
6.1.1 Специальные правовые нормы трудового законодательства.....	49
6.1.2 Эргономические требования к рабочему месту оператора ПЭВМ.....	52
6.2 Производственная безопасность.....	53
6.2.1 Анализ вредных и опасных факторов, которые может создать объект исследования.....	53
6.2.2 Обоснование мероприятий по снижению воздействия вредных и опасных факторов .....	54
6.2.2.1 Микроклимат .....	54
6.2.2.2 Уровень шума.....	56
6.2.2.3 Отсутствие или недостаток естественного света.....	57
6.2.2.4 Повышенная напряженность магнитного поля.....	62
6.2.2.5 Психофизиологический фактор.....	62
6.2.2.6 Электрический ток .....	63

6.3 Экологическая безопасность.....	64
6.4 Безопасность в чрезвычайных ситуациях.....	66
6.4.1 Анализ вероятных ЧС, которые может инициировать объект исследований .....	66
6.4.2 Меры по предотвращению и ликвидации ЧС и их последствий .....	67
6.5 Выводы .....	67
Заключение .....	69
Список научных достижений и публикаций .....	70
Список использованных источников и литературы .....	71
Приложение А .....	74
Приложение Б – Пример документа «Осмотр пациента лечащим врачом» .....	88
Приложение В – Пример документа «Температурный лист» .....	90

## **Введение**

Многие сферы деятельности человека претерпевают различного рода изменения, это связано с совершенствованием, оптимизацией и автоматизацией определенных процессов. Для медицинских учреждений разрабатываются информационные системы, которые позволяют ускорить ввод информации и формирование на их основе различных отчетов и документов. В виду широкого развития аналитики данных и машинного обучения, появляется возможность извлекать полезную информацию из собранных данных, выявлять определенные закономерности.

Данные могут быть представлены в различных форматах: таблицы, изображения, аудио, видео, текст и т.д. Для каждого типа представления данных существует свой подход к обработке и анализу данных. Чаще встречающейся формой представления медицинских данных являются изображения и текст на естественном языке.

Отделением инфекционных заболеваний Сибирского государственного медицинского университета были предоставлены депersonализированные истории болезни пациентов, страдающих рожистыми воспалениями, а также температурные листы, в которых содержится информация о результатах ежедневного измерения температуры, давления и пульса пациентов в период нахождения в медицинском учреждении. История болезни включает в себя документ «Осмотр пациента лечащим врачом», который содержит подробную информацию о состоянии пациента при поступлении в стационар.

Целью работы является повышение эффективности работы врача инфекциониста при анализе данных историй болезней пациентов с бактериологическими инфекциями путем разработки программы выделения значимых диагностических признаков.

В рамках данной работы проводится обработка и подготовка текстовых медицинских данных к анализу. Также предложен подход классификации степеней тяжести заболевания пациентов, страдающих рожистым воспалением.

Объектом исследования являются медицинские документы, сформированные на основе проведенных медицинских осмотров. Предметом исследования является процесс разработки алгоритмического и программного обеспечения анализа результатов медицинских осмотров.

Методы исследования – поиск литературы и источников, анализ информационных материалов, сравнение, консультация со специалистами.

В работе использованы различные методические материалы и интернет-ресурсы. Работа будет реализована на языке программирования «Python».

## 1 Обзор литературы

Широкое распространение понятие «Большие данные» получило относительно недавно, благодаря редактору журнала «Nature», Клиффорду Линчу, в 2008 году. Через пять лет в Оксфордском английском словаре появилось следующее определение данного понятия: «Очень большие массивы данных, которые могут быть проанализированы с помощью компьютеров, чтобы выявить закономерности, тенденции и взаимосвязи, в особенности в отношении поведения людей и их взаимодействий» [1].

Многие сферы человеческой деятельности уже столкнулись с данным понятием. Медицинские, образовательные, финансовые и другие учреждения генерируют большие массивы информации. К сожалению, пока не все способны правильно работать с такими данными, в большинстве случаев данные не сохраняются или же хранятся без применения.

Международной консалтинговой компанией McKinsey было выделено 11 методов и техник, применимых к большим данным, к которым, в частности, относятся машинное обучение и визуализация аналитических данных [2]. Под машинным обучением понимают построение алгоритмов, способных обучаться, средствами математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов и различных техник работы с цифровыми данными. Визуализация аналитических данных – это представление исходной информации и результатов анализа данных в наиболее удобной для восприятия и интерпретации форме.

Помимо больших размеров, особенностью больших данных является способ их представления. В большинстве случаев, данные представляются в неструктурированном или слабоструктурированном виде. Неструктурированный способ представления подразумевает отсутствие определенной структуры, слабоструктурированный – не соответствует строгой структуре, однако, содержит теги и другие маркеры для отделения семантических элементов.

Многие медицинские учреждения для автоматизации своей деятельности разрабатывают информационные системы, функционал которых ограничивается вводом информации и формированием на их основе документов и отчетов. Расширить возможности медицинских информационных систем можно с помощью методов и техник, применимых к большим данным.

Медицинские документы, в большинстве случаев, представляют собой текстовую информацию, написанную на естественном языке, хранимую на бумажном или электронном носителе.

Работа с текстовыми данными имеет свои особенности. Текст, который читает человек, машиной воспринимается иначе. Единицей анализа текста является не слово, а лексема, т.е. строка кодированных байтов, представляющих текст. Строка кодированных байтов не отражает семантику слова, не указывает на связи и зависимости слов. Поэтому без предварительной подготовки текста применение методов машинного обучения не является корректным. К этапам подготовки текста для анализа относят токенизацию, стемминг (лемматизацию), исключение стоп-слов, а также кодирование слов в числовые значения. Под токенизацией понимается разбиение текста на более мелкие части, токены, к ним относят слова и знаки пунктуации [3]. Стемминг (лемматизация) – это процесс нахождения лексической основы для заданного исходного слова с применением основных правил языка. Стоп-слова – это слова, которые не несут никакой смысловой нагрузки.

В настоящее время существует большое количество методов, обеспечивающих форматирование текстовых данных к виду, пригодному для подачи на вход алгоритмам машинного обучения. Самым простым способом кодировки слов является их нумерация, однако, числовые данные, полученные в результате данного способа, не несут никакой смысловой нагрузки. По номерам слов невозможно определить насколько близкими по смыслу они являются друг другу.

Существует также способ отражения наличия или отсутствия слова в тексте, называемый «One-hot encoding». Суть данного метода заключается в

том, что создается матрица, столбцами которой являются слова, а строками являются документы. В каждой конкретной ячейке данной матрицы записывается либо 0, либо количество встречаемости слова в документе [4].

Повторение одного и того же слова в нескольких документах свидетельствует о том, что данное слово не может быть важным и не несет много информации. Для решения данной проблемы определяется вес слова с помощью статистической меры, называемой «TF-IDF» [5]. Большой вес в TF-IDF получают слова с высокой частотой в пределах конкретного документа и с низкой частотой употреблений в других документах. Данная мера часто используется в задачах информационного поиска и анализа текстов.

Для данной работы наиболее подходящим является метод «TF-IDF», так как существует необходимость в оценке важности слов в контексте не обусловленной лишь частой употреблением слова.

## **2 Объект и методы исследования**

### **2.1 Объект исследования**

Объектом исследования данной работы являются медицинские документы, сформированные на основе проведенных медицинских осмотров. Предметом исследования является процесс разработки алгоритмического и программного обеспечения анализа и визуализации результатов медицинских осмотров.

Отделением инфекционных заболеваний Сибирского государственного медицинского университета были предоставлены деперсонализированные истории болезни пациентов, страдающих рожистым воспалением, а также температурные листы, в которых хранится информация об ежедневных измерениях температуры, давления и пульса пациентов в период нахождения в медицинском учреждении. Документ «Осмотр пациента лечащим врачом» является частью истории болезни и состоит из следующих основных 14 блоков информации:

1. Номер пациента, пол и возраст;
2. Дата и время осмотра;
3. Жалобы (описание факторов, беспокоящих пациента, с его слов);
4. Анамнез болезни (описание этапов развития симптомов заболевания до поступления в медицинское учреждение, со слов пациента);
5. Эпиданамнез (описание сведений о контактах пациента с инфекционными больными, о поездках в регионы эндемичные по определенной инфекции, о приеме в пищу определенных продуктов);
6. Анамнез жизни (описание условий жизни и труда пациента, его пищевых и поведенческих привычек, а также ранее перенесенных заболеваний);
7. Анамнез врачебно-трудовой экспертизы (информация о листе нетрудоспособности);
8. Объективный статус (описание наличия или отсутствия патологий по каждому из рассматриваемых органов и систем организма);

9. Локальный статус (описание исследования пораженной системы);
10. Диагноз при поступлении;
11. Обоснование диагноза (описание обоснования диагноза с указанием данных, которые подтверждают указанный диагноз);
12. Диагноз сопутствующий (описание наличия или отсутствия сопутствующих заболеваний);
13. План обследования (описание необходимых этапов обследования);
14. План лечения (описание медицинских препаратов и действий, которые будут применяться во время лечения).

Для каждого отдельного пациента количество блоков информации документа «Осмотр лечащим врачом» варьируется. Документы «Осмотр пациента лечащим врачом» и «Температурный лист» имеются по 36 пациентам. В приложениях Б и В представлены примеры документов «Осмотр лечащим врачом» и «Температурный лист», соответственно.

Одним из принципов клинической классификации рожистого воспаления является степень тяжести [6]. По степени тяжести выделяют легкую (I), среднетяжелую (II) и тяжелую (III) формы заболевания. Форма заболевания, выявленная при поступлении пациента в медицинское учреждение, определяет ход действий, которые необходимо производить над пациентом в дальнейшем. В блоке «Диагноз» документа «Осмотр пациента лечащим врачом» указывается степень тяжести заболевания, однако более достоверным является определение степени тяжести заболевания с помощью параметров, представленных в таблице 1.

Таблица 1 – Параметры для определения степени тяжести рожистого воспаления

Параметр \ Степень тяжести	I	II	III
Температура	< 38	38 - 40	> 40
Длительность лихорадочного периода	1-2 дня	3-4 дня	> 4 дней
Инттоксикация	Слабо выражена	Умеренно выражена	Сильно выражена
Головная боль	-	+	Сильная
Озноб	-	+	-

## Продолжение таблицы 1

Параметр \ Степень тяжести	I	II	III
Мышечные боли	-	+	-
Пониженное давление	-	+	+
Тошнота, рвота	-	+	+
Учащенный пульс	-	+	Сильно
Адинамия	-	-	+
Бред, спутанность сознания	-	-	+
Судороги	-	-	+
Местный процесс	Локализованный, эритематозный	Локализованный, распространенный	Выраженный, распространенный

## 2.2 Методы исследования

### 2.2.1 Метод определения нормальной формы слова

При анализе текстовых данных различные формы одного слова воспринимаются как разные слова. По этой причине возникает необходимость в поиске нормальной формы слов. Под нормальной формой слова, с лингвистической точки зрения, понимается начальная форма слова. В русском языке нормальными считаются следующие морфологические формы слов:

- существительные в именительном падеже и единственном числе;
- прилагательные в именительном падеже, единственном числе и мужском роде;
- глаголы, причастия и деепричастия в форме инфинитива глагола несовершенной формы.

Наиболее результативным методом определения нормальной формы слова является лемматизация, т.е. процесс приведения слова к лемме – ее нормальной (словарной) форме. По сравнению со стеммингом, при лемматизации не удаляются части слова, окончания слов подвергаются замене на окончания слов начальной формы. При работе лемматизатора используется словарь, из которого извлекаются необходимые формы слов. После чего исходные слова заменяются на их нормальные формы.

### 2.2.2 Метод определения веса признаков документа «TF-IDF»

По завершении формирования списка однозначно определенных несущих смысловую нагрузку слов, необходимо произвести форматирование слов в числовые значения. Наиболее распространенными являются такие

методы, как замена слов на их порядковый номер в списке всех существующих слов в документе или же создание матрицы на пересечении строк (документов) и столбцов (слов) которой указываются 0 или 1, свидетельствующие об отсутствии или наличии слов в документах, соответственно. Последний метод получил название «OneHotEncoder», наиболее информативной его доработкой является указание на пересечении строк и столбцов не наличия или отсутствия слова в документе, а его веса. Повторение одного и того же слова в нескольких документах свидетельствует о том, что данное слово не является достаточно важным и не несет много информации. Для определения веса слова используется статистическая мера, называемая «TF-IDF» [5].

Для каждого слова рассчитывается вес, оценивается важность слова в пределах отдельного документа. Оценка важности слова  $t_i$  в пределах отдельного документа рассчитывается по следующей формуле:

$$tf(t, d) = \frac{n_t}{\sum_k n_k}, \quad (1)$$

где  $n_t$  - число вхождений слова  $t$  в документ, а в знаменателе – общее число слов в данном документе.

Формула для расчета инверсии частоты, с которой некоторое слово встречается в документах коллекции, следующая:

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D | t \in d_i\}|}, \quad (2)$$

где  $|D|$  – число документов в коллекции,  $|\{d_i \in D | t \in d_i\}|$  - число документов из коллекции  $D$ , в которых встречается  $t$  (когда  $n_t \neq 0$ ).

Выбор основания логарифма в формуле не имеет значения, поскольку изменение основания приводит к изменению веса каждого слова на постоянный множитель, что не влияет на соотношение весов. Мера «TF-IDF» является произведением следующих двух сомножителей:

$$tf - idf(t, d, D) = tf(t, d) \times idf(t, D) \quad (3)$$

Большой вес в TF-IDF получают слова с высокой частотой в пределах конкретного документа и с низкой частотой употреблений в других

документах. Данная мера часто используется в задачах информационного поиска и анализа текстов.

### 2.2.3 Метод дерева решений

Дерево решений является одним из наиболее часто и широко используемых алгоритмов машинного обучения с учителем. Данный метод применим как к задачам классификации, так и регрессии.

Для каждого атрибута в наборе данных алгоритм дерева решений формирует узел, самый важный атрибут размещается в корневом узле. Двигаясь от корневого узла, происходит спуск по дереву, следуя соответствующему узлу, который соответствует условию или «решению». Данный процесс продолжается до тех пор, пока не будет достигнут последний узел, который содержит прогноз или результат дерева решений.

Для лучшего понимания данного метода, рассмотрим сценарий, когда человек просит одолжить ему машину на один день, необходимо принять решение, одолжить ему машину или нет. Дерево решений для данного сценария представлено на рисунке 1.

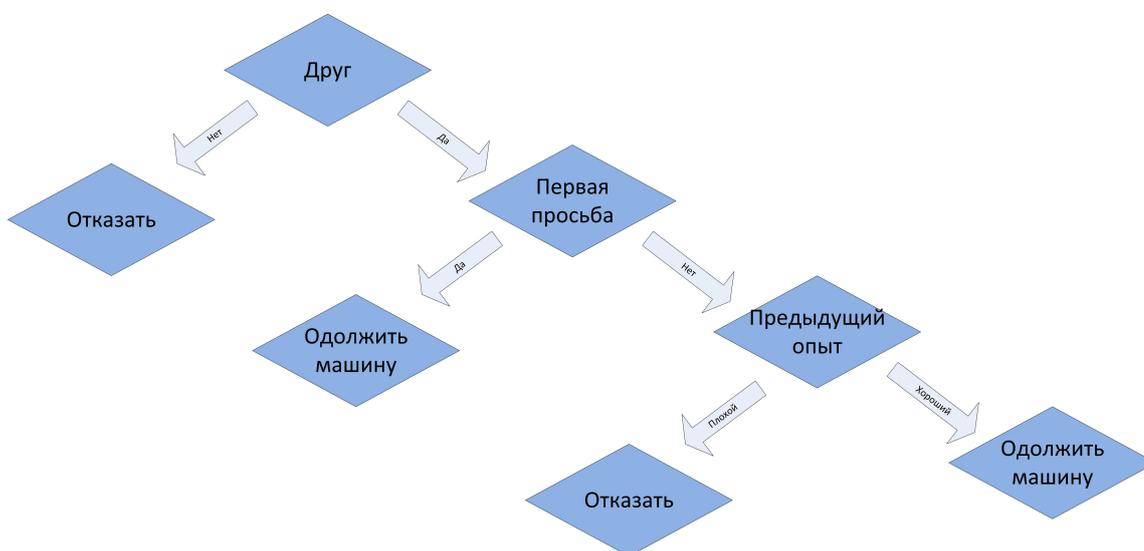


Рисунок 1 – Дерево решений «Одолжить ли машину?»

Есть несколько факторов, которые помогают определить решение, некоторые из которых перечислены ниже:

1. Человек является другом или просто знакомый? Если человек просто знакомый, то просьба будет отклонена; если человек друг, то переход к следующему шагу.

2. Человек спрашивает машину в первый раз? Если это так, то можно одолжить машину, в противном случае переход к следующему шагу.

3. Была ли машина повреждена в последний раз, когда он возвращал ее? Если да, то просьба будет отклонена; если нет, то можно одолжить машину.

### 2.2.4 Метод градиентного бустинга

Градиентный бустинг – это техника для выполнения задач машинного обучения с учителем, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Принцип действия данного метода представлен на рисунке 2.

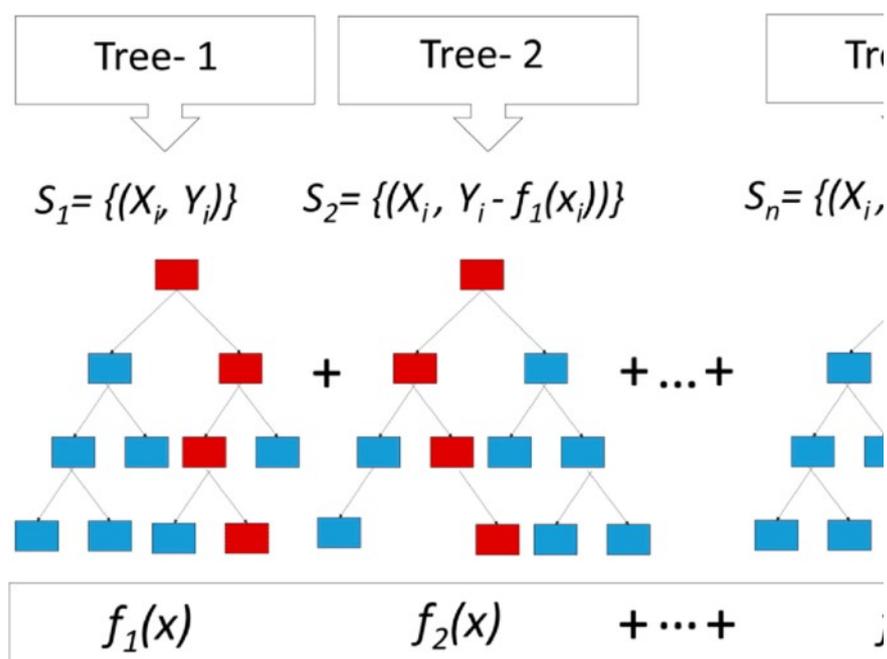


Рисунок 2 – Принцип работы градиентного бустинга

Бустинг строит модели из отдельных «слабых учеников» итеративным способом. Отдельные модели строятся не на случайных подмножествах данных и функций, а на согласованной основе, придавая больший вес экземплярам с неправильными предсказаниями и высокими ошибками. Общая идея заключается в том, что случаи, которые трудно предсказать правильно («сложные» случаи), будут сконцентрированы в процессе обучения, так что модель учится на ошибках прошлого. Тренировку каждого ансамбля на

подмножестве обучающего набора также называют стохастическим градиентным бустингом, который может помочь улучшить обобщаемость модели.

Градиент используется для минимизации функции потерь. На каждом этапе обучения создается слабый ученик, и его прогнозы сравниваются с правильным результатом, который является ожидаемым. Расстояние между предсказанием и истиной представляет частоту ошибок модели. Эти ошибки используются для расчета градиента. Градиент представляет собой в основном частную производную функции потерь - поэтому она описывает крутизну функции ошибок. Градиент можно использовать для определения направления, в котором нужно изменить параметры модели, чтобы (максимизировать) уменьшить ошибку в следующем раунде обучения, «снижая градиент».

### **3 Расчеты и аналитика**

#### **3.1 Выбор программного обеспечения**

Среди языков программирования по работе с данными лидирующие позиции занимают такие языки, как R и Python [7]. Оба языка обладают открытым исходным кодом, являются платформенно независимыми, необходимые пакеты легко устанавливаются, настраиваются и применяются.

В R помимо анализа данных реализуется визуализация данных, в чем Python, обладающий множеством готовых библиотек, не уступает.

R является менее популярным, поэтому уступает Python в количестве массивов документации, требуемой разработчикам для создания приложений. К недостаткам Python относят сложность отслеживания ошибок в коде.

Что касается простоты использования, то Python можно отнести к той группе языков, которые подходят на роль первого языка программирования, в то время как R является узкоспециализированным языком.

Преимуществом Python также являются особенности синтаксиса, стимулирующие разработчика писать удобно читаемый код, при использовании R многие пользователи пренебрегают рекомендациями по оформлению программного кода, что ухудшает его читабельность.

В результате анализа преимуществ и недостатков R и Python, было принято решение, что наиболее подходящим инструментом для решения поставленной задачи, т.е. разработки алгоритма классификации, а также визуализации результатов работы алгоритма, является Python.

#### **3.2 Загрузка и предварительный анализ данных**

Исходные данные хранятся в файлах с расширениями «.txt» и «.xls» по каждому пациенту отдельно. Текстовые файлы содержат в себе документы «Осмотр лечащим врачом», а Excel-файлы содержат информацию об измерениях температуры, давления и пульса. Примеры документов «Осмотр лечащим врачом» и «Температурный лист» представлены в приложениях Б и В, соответственно.

В объект «list» заносятся текстовые данные по каждому отдельному пациенту. С помощью библиотеки «pandas» формируется объект «DataFrame», в который считываются транспонированные таблицы из файлов с расширением «.xls».

Элементы «DataFrame» подвергаются предварительному анализу: проверка на выбросы и пропущенные значения. На рисунке 3 представлена визуализация результатов анализа набора данных на наличие пропущенных значений.

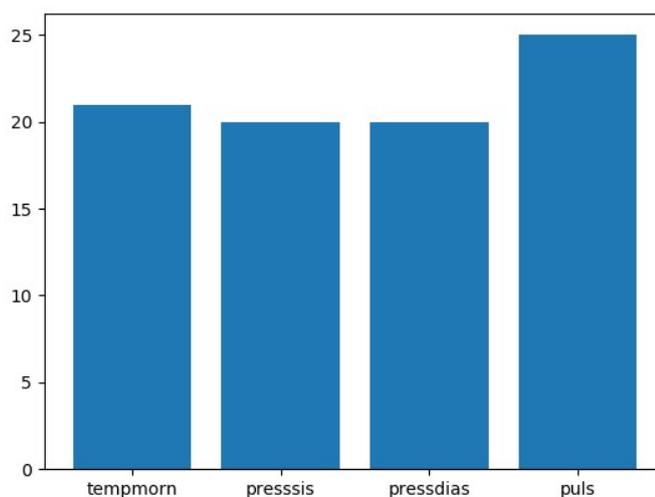


Рисунок 3 – Наличие пропущенных значений

Поиск пропущенных значений осуществляется с помощью «df.isnull().sum()». Так, пропущенные значения содержат такие показатели как температура утром, систолическое давление, диастолическое давление и пульс. Существует несколько способов по устранению пропущенных значений. Строки или столбцы, содержащие пропущенные значения, можно удалить, а также пропущенные значения можно импутировать, т.е. заменить на среднее, моду, медиану по столбцу или строке или константу. Вариант с удалением строк, столбцов в данном случае не подходит, поэтому было решено произвести импутацию пропущенных значений на среднее по столбцу «SimpleImputer(missing\_values=np.nan, strategy='mean')».

Затем проверили поля на наличие выбросов. Под выбросами понимают такие значения показателей, которые значительно отличаются от основного множества значений. На рисунках 4, 5 и 6, представлены визуализации

результатов анализа на наличие выбросов таких показателей, как температура утром и вечером, давление систолическое и диастолическое и пульс, соответственно.

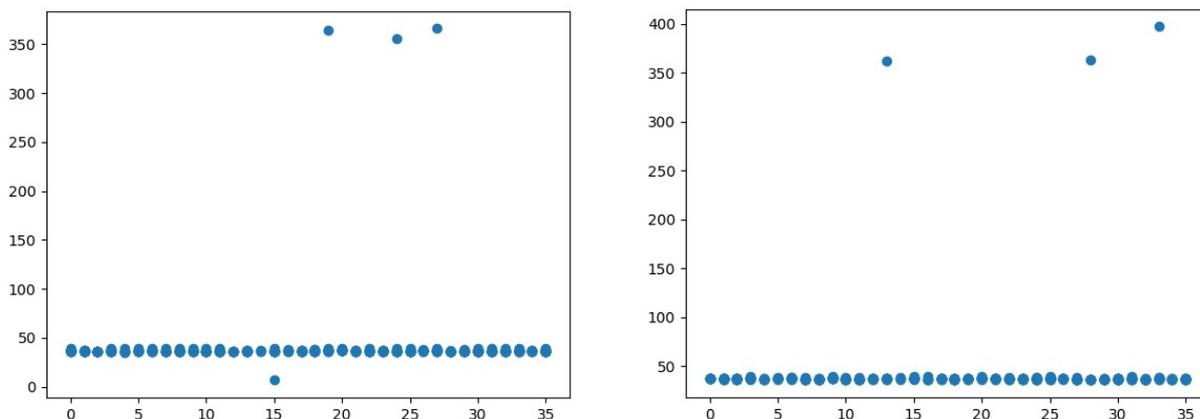


Рисунок 4 – Значения показателей «Утренняя температура» и «Вечерняя температура»

Из графиков можем увидеть, что основная часть значений утренней и вечерней температуры находится в диапазоне ниже 50 и выше 0, однако, четыре значения на графике слева и три значения на графике справа значительно от них отличаются, т.е. данные точки можно считать выбросами.

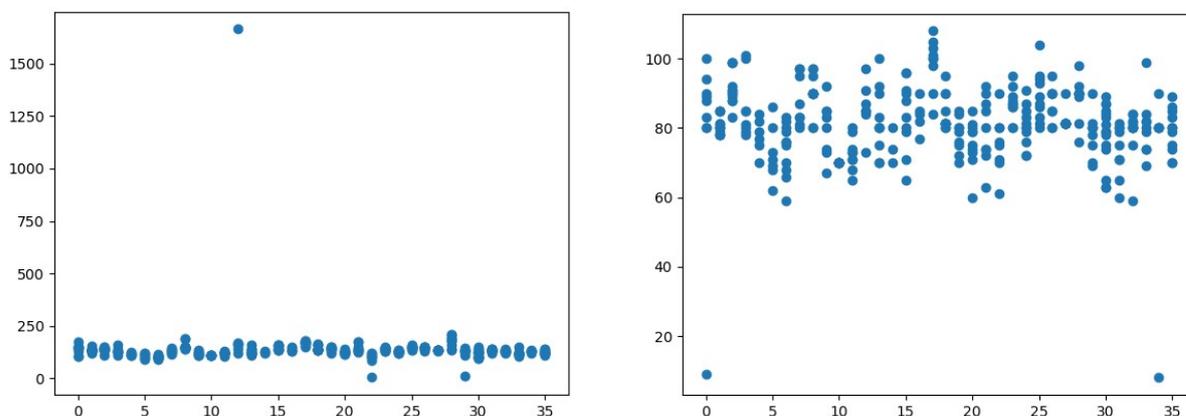


Рисунок 5 – Значения показателей «Систолическое давление» и «Диастолическое давление»

На основе данных графиков можем сделать вывод, что показатели «Систолическое давление» и «Диастолическое давление» имеют три и два выброса, соответственно.

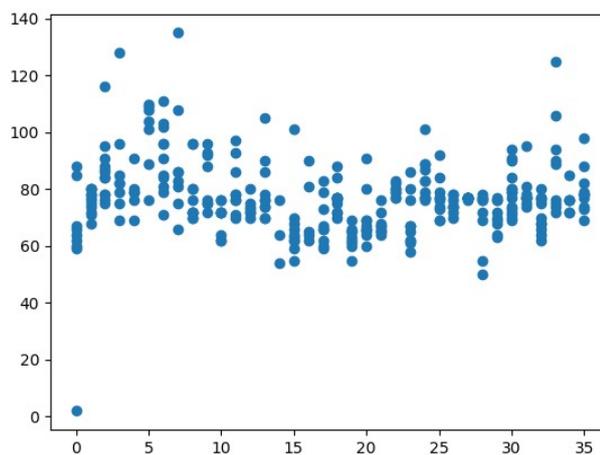


Рисунок 6 – Значения показателя «Пульс»

Такой показатель как «Пульс» имеет только одно значение, которое можно считать выбросом. Наличие выбросов оказывает влияние на результаты дальнейшей работы, поэтому было решено выбросы, так же как и пропущенные значения, заменить на среднее значение по столбцу.

### 3.3 Предварительная подготовка, построение «TF-IDF» матрицы и определение значений меток классов

Элементы списка подвергаются обработке. Все символы переводятся в нижний регистр, затем формируется список стоп-слов для русского языка с помощью «`nltk.corpus.stopwords.words('russian')`». Элементы списка текстовых данных токенизируются с помощью «`yargy.tokenizer.MorphTokenizer()`». Среди полученных токенов удаляются стоп-слова. Затем производится лемматизация оставшихся токенов с помощью метода «`normalized`».

В виду того, что алгоритмы машинного обучения предпочтительно принимают на вход символы латиницы, все слова были подвержены транслитерации. Для этого был создан словарь, ключами которого являются символы русского алфавита, а значениями – представление данных букв латинскими символами.

По завершении предобработки текста строится «TF-IDF» матрица с помощью «`sklearn.feature_extraction.text.TfidfVectorizer`». Затем объединяются два объекта «`DataFrame`»: с данными из Excel-файлов и «TF-IDF» матрицы, после чего определяются значения меток классов на основе таких показателей

как температура, давление (ниже нормы), пульс (учащенный), головная боль, озноб, мышечные боли, тошнота, рвота, бред, спутанность сознания, судороги, местный процесс. После определения значений меток классов перечисленные показатели удаляются из общего набора данных.

### **3.4 Разделение набора данных на тестовую и тренировочную выборки**

Для обеспечения проверки качества работы модели классификации возникает необходимость в разделении общего набора данных на тренировочную и тестовую выборки. Тренировочная выборка предназначена для обучения модели, тестовая выборка – для оценки качества работы модели.

Для разбиения набора данных на тренировочную и тестовую выборки используется метод «`sklearn.model_selection.train_test_split`», в качестве параметра было указано, что тестовая выборка будет составлять 30% от общего набора данных.

## **3.5 Построение классификатора**

### **3.5.1 Выбор модели классификации**

В ходе работы для определения наилучшей модели классификации были рассмотрены следующие модели:

- «Дерево решений» – «`sklearn.tree.DecisionTreeClassifier`»;
- «Случайный лес» – «`sklearn.ensemble.RandomForestClassifier`»;
- «Градиентный бустинг» – «`xgboost.sklearn.XGBClassifier`».

Перечисленные классификаторы рассматриваются без специально подобранных параметров. Каждая модель оценивается по десять раз на случайном небольшом наборе записей из набора данных. С помощью метрики точности «ассигасу» оценивается качество модели. Модель выбирается на основе наибольшего среднего значения точности десяти моделей этого типа. Для текущей задачи метод «Градиентный бустинг» стал наилучшей моделью.

### 3.5.2 Подбор оптимальных параметров классификатора

После выбора модели было решено подобрать параметры, которые определяют ее эффективность. Алгоритм «XGBClassifier» может принимать следующие параметры:

- «max\_depth» (по умолчанию = 6) – максимальная глубина дерева;
- «learning\_rate» (по умолчанию = 0,3) – размер шага, используемый при обновлении, предотвращает переоснащение;
- «objective» – задача обучения и соответствующая цель обучения или специальная целевая функция, которая будет использоваться;
- «booster» – используемый бустер: gbtree, gblinear или dart;
- «gamma» (по умолчанию = 0) – минимальное уменьшение потерь, необходимое для создания дальнейшего разбиения на листовом узле дерева;
- «min\_child\_weight» (по умолчанию = 1) – минимальная сумма веса экземпляра, необходимая для потомка, если в результате шага разбиения дерева получится листовой узел с суммой веса экземпляра меньше данного параметра, то дальнейшее разбиение прекратится;
- «subsample» (по умолчанию = 1) – соотношение подвыборок учебных экземпляров;
- «colsample\_bytree» (по умолчанию = 1) – отношение выборки столбцов при построении каждого дерева, подвыборка выполняется один раз для каждого построенного дерева;
- «scale\_pos\_weight» (по умолчанию = 1) – балансировка положительных и отрицательных весов.

Было решено установить следующие значения для параметров: «objective» = binary:logistic, «booster» = gbtree, «gamma» = 0, «min\_child\_weight» = 1, «scale\_pos\_weight» = 1. Остальные параметры будем варьировать следующим образом при подборе с помощью перекрестной проверки:

«max\_depth» = [3, 4, 5, 6]

«learning\_rate» = [0,1, 0,2, 0,3, 0,4]

«subsample» = [0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1]

«colsample\_bytree» [0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1]

Каждая комбинация будет проверяться с использованием кросс-проверки. Данная проверка позволяет выбрать лучшую комбинацию параметров. Итоговая модель затем обучится на всем множестве с использованием выбранных параметров.

С помощью «sklearn.model\_selection.GridSearchCV» проведем исчерпывающий поиск по сетке параметров. В качестве параметров укажем метрику качества – точность (accuracy) и пятикратную перекрестную проверку.

### **3.5.3 Построение классификатора с оптимальными параметрами**

После подбора оптимальных параметров построим модель на обучающем подмножестве данных, а затем предскажем значения на тестовом наборе.

На основе значений меток для тестового набора и предсказанных значений оценим качество работы модели. Построим матрицу ошибок с помощью «sklearn.metrics.confusion\_matrix» и визуализируем ее. Элементы матрицы, лежащие на главной диагонали показывают количество верно определенных экземпляров, а недиагональные элементы свидетельствуют о количестве экземпляров тестового подмножества отнесенных к тому или иному классу не верно. Также для оценки качества модели используются такие показатели, как точность, полнота и верность.

## 4 Результаты

### 4.1 Классификатор степеней тяжести рожистого воспаления

Со значениями параметров по умолчанию были построены три модели классификации степеней тяжести заболевания. Результаты работы таких моделей, как «Дерево решений», «Случайный лес» и «Градиентный бустинг», а именно, значения точности их предсказаний, представлены в таблице 2.

Таблица 2 – Точность предсказания моделей классификации

№	Дерево решений	Случайный лес	Градиентный бустинг
1	0,94	0,96	0,94
2	0,89	0,86	0,89
3	0,94	0,96	0,96
4	0,89	0,91	0,89
5	0,95	0,94	0,95
6	0,93	0,91	0,93
7	0,93	0,90	0,93
8	0,97	0,93	0,97
9	0,94	0,96	0,94
10	0,95	0,97	0,95
Среднее	0,933	0,93	0,935

Наибольшей средней точностью обладает модель «Градиентный бустинг». Затем для данной модели был проведен подбор следующих оптимальных параметров:

```
XGBClassifier(booster='gbtree', colsample_bytree=0.1, gamma=0, learning_rate=0.1, max_depth=3, min_child_weight=1, n_estimators=100, objective='binary:logistic', scale_pos_weight=1, subsample=0.3, tree_method='exact')
```

Для оценки построенной модели была визуализирована матрица ошибок, представленная на рисунке 7.

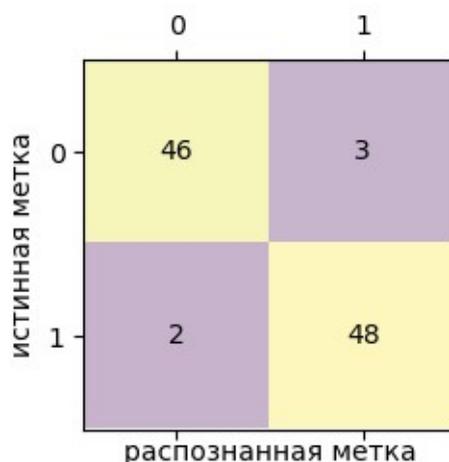


Рисунок 7 – Матрица ошибок

Из матрицы ошибок видно, что модель классифицирует большую часть выборки верно. Так, 46 записей о пациентах со средней степенью тяжести заболевания и 48 записей с тяжелой степенью отнесены верно и только два тяжелых случая и три средней формы классифицированы неверно.

Значения таких метрик качества как верность (accuracy), полнота (recall), точность (precision) и F1 представлены в таблице 3.

Таблица 3 – Значения метрик качества работы классификатора

Метрика	Верность	Полнота	Точность	F1
Значение	0,95	0,96	0,94	0,95

Доля правильных прогнозов составила 95%, доля истинно положительных исходов составила 96%. Значения всех метрик достаточно высоки, что говорит о высоком качестве работы алгоритма классификации.

#### 4.2 Определение важности признаков

При использовании градиентного бустинга, есть возможность в определении наиболее важных признаков. На рисунке 8 представлена визуализация важности признаков.

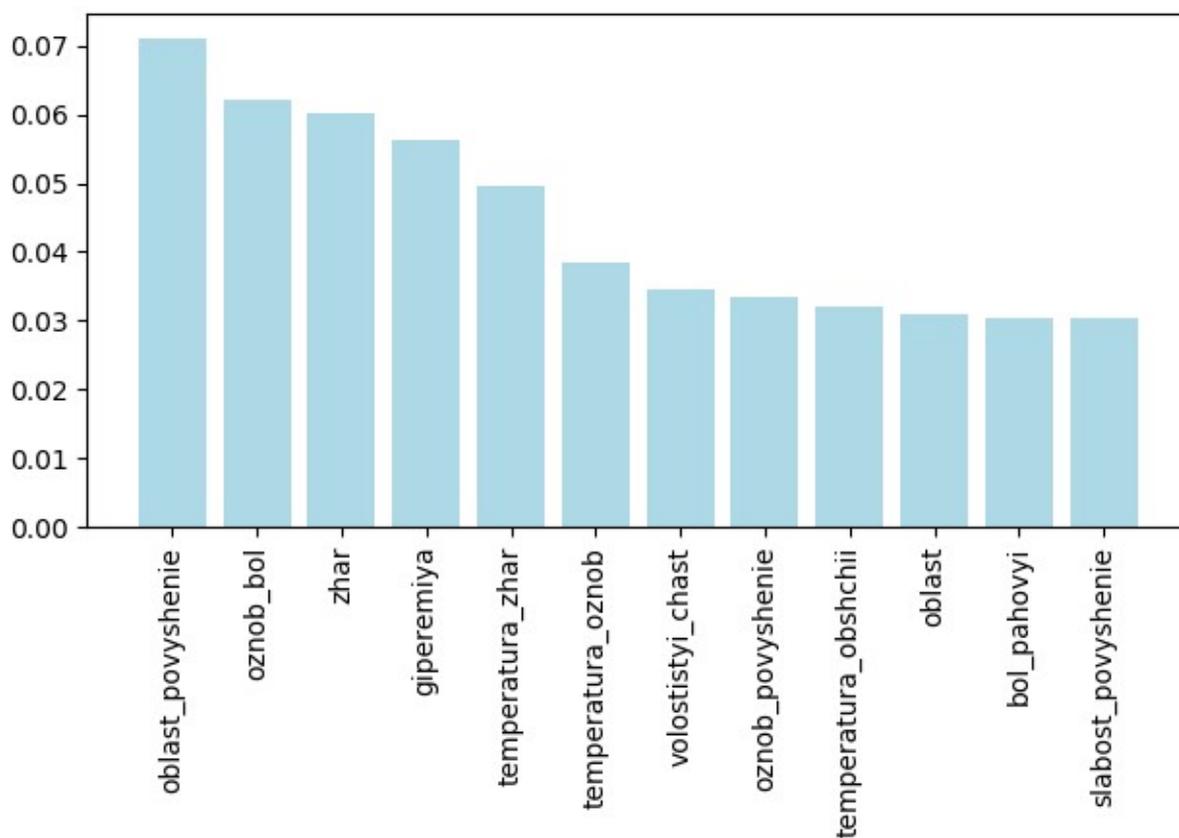


Рисунок 8 – Важности признаков классификации

Можем сделать вывод, что наиболее важным для классификации степеней тяжести рожистых воспалений является наличие таких слов и словосочетаний, как «область\_повышение», «озноб\_боль», «жар», «гиперемия» и «температура\_жар».

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

Студенту:

<b>Группа</b>	<b>ФИО</b>
8ПМ8И	Кашеевой Евгении Викторовне

<b>Инженерная школа</b>	<b>ИШИТР</b>	<b>Отделение (НОЦ)</b>	Информационных технологий
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.09.04 Программная инженерия

Тема ВКР:

Алгоритмическое и программное обеспечение анализа и визуализации результатов медицинских осмотров

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

<i>1. Стоимость ресурсов научно-исследовательского проекта: материально-технических, энергетических, финансовых, информационных и человеческих</i>	Материальные затраты – 1658 руб.; Затраты на электроэнергию, мобильную связь и интернет – 1078,32 руб.; Амортизационные затраты - 4366,88; Оклад научного руководителя - 33664 руб.; Оклад инженера- 21760 руб.
<i>2. Нормы и нормативы расходования ресурсов</i>	Норма амортизации – 2,8; Районный коэффициент – 1,3.
<i>3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Коэффициент отчислений на уплату во внебюджетные фонды 30 %.

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

<i>1. Оценка коммерческого и инновационного потенциала научно-исследовательского проекта</i>	Анализ перспективности технических решений посредством QuaD-анализа Диаграмма Исикавы Оценка готовности научно-исследовательского проекта к коммерциализации
<i>2. Разработка устава научно-исследовательского проекта</i>	Определение цели научно-исследовательского проекта, требований к проекту, описание заинтересованных сторон проекта, рабочей группы.
<i>3. Планирование процесса управления научно-исследовательским проектом: структура и график проведения, бюджет, риски и организация закупок</i>	Планирование этапов работы, определение календарного графика проведения исследования. Определение рисков научно-исследовательского проекта, оценка вероятности риска и потерь. Расчет бюджета затрат на проведение исследования.
<i>4. Определение ресурсной, финансовой, экономической эффективности</i>	Описание потенциального эффекта научно-исследовательского проекта.

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. Оценочная карта технологии «QuaD»
2. Диаграмма Исикавы
3. Диаграмма Ганта
4. Бюджет затрат
5. Реестр рисков

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

**Задание выдал консультант:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент	Меньшикова Екатерина Валентиновна	к.ф.н., доцент		

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8ПМ8И	Кащеева Евгения Викторовна		

## **5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение**

Данная работа посвящена разработке алгоритма классификации степеней тяжести заболевания по данным из историй болезни пациентов с бактериологическими инфекциями. Данный алгоритм может найти свое применение в системах поддержки принятия решений врача. Подобные системы смогут помочь врачу сократить время, а также увеличить точность принятия решений о способе лечения пациентов на основе схожих случаев.

В качестве потенциальных потребителей результатов научно-исследовательского проекта могут выступать организации здравоохранения, медицинский персонал, а также представители научного сообщества, заинтересованные в продолжении исследования или в использовании его технических результатов.

С целью эффективного использования научного потенциала проекта необходимо уделять внимание, как разработке, так и проведению её анализа с точки зрения экономических требований.

В качестве задач данного раздела выступают следующие:

- Определить перспективность научно-исследовательского проекта с помощью технологии «QuaD»;
- Проанализировать причины низкого уровня автоматизации деятельности врача, используя диаграмму Исикавы;
- Оценить готовность научно-исследовательского проекта к коммерциализации;
- Провести инициацию научно-исследовательского проекта;
- Составить план научно-исследовательского проекта, провести планирование бюджета;
- Провести анализ рисков.

## 5.1 Предпроектный анализ

### 5.1.1 Технология «QuaD»

В виду того, что медицинские информационные системы разрабатываются и используются исключительно в рамках конкретного медицинского учреждения и не распространяются за его пределы, анализ конкурентных технических решений невозможен ввиду отсутствия открытых данных о наличии и свойствах подобных решений.

Одним из инструментов измерения характеристик качества новой разработки и ее перспективности на рынке является технология «QuaD» [8], которая позволяет принимать решение о целесообразности вложения денежных средств в научно-исследовательский проект. В основе данной технологии лежит нахождение средневзвешенного значения показателя качества и перспективности научной разработки  $P_{cp}$ .

$$P_{cp} = B_i \cdot B_i, \quad (4)$$

где  $B_i$  – вес показателя;  $B_i$  – средневзвешенного значение показателя.

Если значение показателя  $P_{cp}$  варьируется в диапазоне от 100 до 80, то разработку принято считать перспективной. В случае, когда данный показатель входит в диапазон от 79 до 60, то перспективность разработки выше среднего. Средняя, ниже среднего и низкая перспективность считается при значениях показателя от 69 до 40, от 39 до 20 и от 19 и ниже, соответственно. Результаты расчетов средневзвешенных значений показателей представлены в таблице 4.

Таблица 4 – Оценочная карта технологии «QuaD»

Критерий оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение	Средневзвешенное значение
Энергоэффективность	0,15	100	100	1	0,15
Помехоустойчивость	0,1	90	100	0,9	0,09
Надежность	0,15	95	100	0,95	0,1425
Унифицированность	0,05	70	100	0,7	0,035
Время выполнения алгоритма	0,05	100	100	1	0,05
Пользовательский интерфейс	0,05	0	100	0	0
Безопасность	0,1	100	100	1	0,1

## Продолжение таблицы 4

Потребность в ресурсах памяти	0,1	75	100	0,75	0,075
Функциональная мощность	0,05	80	100	0,8	0,04
Простота эксплуатации	0,1	100	100	1	0,1
Качество интеллектуального интерфейса	0,05	80	100	0,8	0,04
Прозрачность кода	0,05	100	100	1	0,05
Итого	1				0,8725

В виду того, что у рассматриваемой разработки наиболее важные критерии имеют высокие показатели, что в сумме составляет 0,8725, разработка является перспективной.

### 5.1.2 Диаграмма Исикавы

Диаграмма причины-следствия Исикавы позволяет графически проанализировать и сформировать причинно-следственные связи, это инструментальное средство для систематического определения причин проблемы и последующего графического представления [9].

Проблемной областью анализа является низкий уровень автоматизации деятельности врача. На рисунке 9 представлена причинно-следственная диаграмма Исикавы.

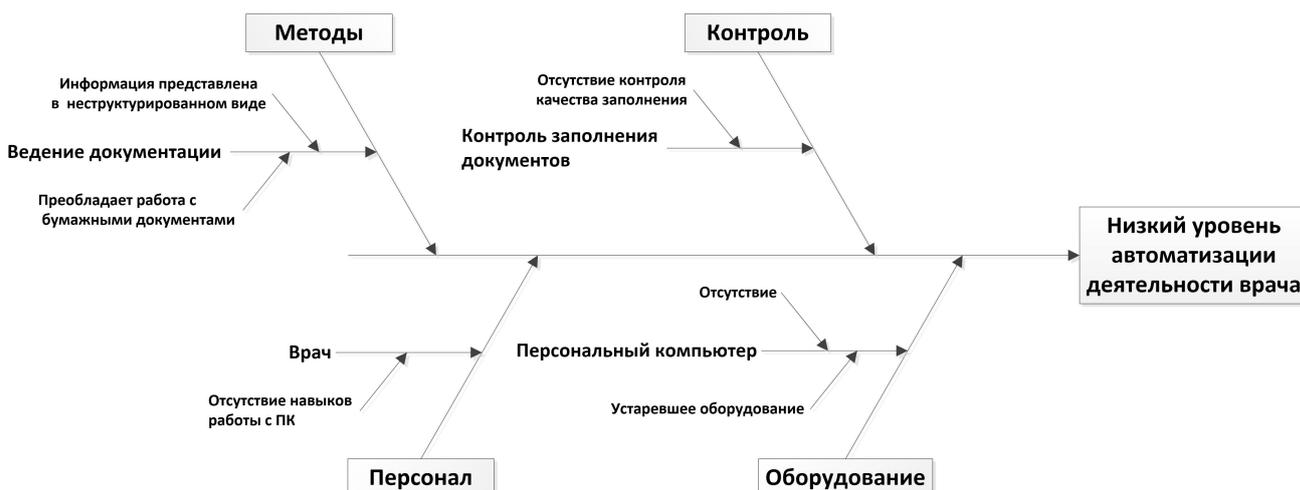


Рисунок 9 – Причинно-следственная диаграмма

Источниками низкого уровня автоматизации деятельности врача могут быть персонал, у которого отсутствуют навыки работы с персональным компьютером, устаревшее оборудование, методы работы с медицинской

документацией, а также отсутствие контроля качества заполнения документации.

### 5.1.3 Оценка готовности научно-исследовательского проекта к коммерциализации

Заполнение бланка оценки готовности научного проекта к коммерциализации позволяет оценить степень готовности разработки к коммерциализации и выяснить уровень собственных знаний для ее применения на любом из этапов ее жизненного цикла.

По каждому из показателей выставляется оценка по пятибалльной шкале, при этом система измерений по каждому из направлений (степень проработанности научного проекта, уровень знаний разработчика) отличается. Система баллов приведена в таблице 5.

Таблица 5 – Система баллов

Балл	Направление	
	Степень проработанности проекта	Уровень знаний разработчика
1	Непроработанность	Не знаком или мало знаю
2	Слабая проработанность	В объеме теоретических знаний
3	Выполнено, но в качестве не уверен	Знаю теорию и практические примеры
4	Выполнено качественно	Знаю теорию и самостоятельно выполняю
5	Имеется положительное заключение независимого эксперта	Знаю теорию, выполняю и могу консультировать

Заполненный бланк приведен в таблице 6.

Таблица 6 – Бланк оценки степени готовности научно-исследовательского проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научно-исследовательского проекта	Уровень имеющихся знаний у разработчика
1	Определен имеющийся научно-технический задел	4	4
2	Определены перспективные направления коммерциализации научно-технического задела	4	4
3	Определены отрасли и технологии (товары, услуги) для предложения на рынке	4	4

Продолжение таблицы 6

4	Определена товарная форма научно-технического задела для представления на рынок	4	4
5	Определены авторы и осуществлена охрана их прав	3	3
6	Проведена оценка стоимости интеллектуальной собственности	2	2
7	Проведены маркетинговые исследования рынков сбыта	2	2
8	Разработан бизнес-план коммерциализации научной разработки	1	1
9	Определены пути продвижения научной разработки на рынок	2	2
10	Разработана стратегия (форма) реализации научной разработки	2	2
11	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	1	1
12	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	1	1
13	Проработаны вопросы финансирования коммерциализации научной разработки	1	1
14	Имеется команда для коммерциализации научной разработки	1	1
15	Проработан механизм реализации научно-исследовательского проекта	3	3
	ИТОГО БАЛЛОВ	35	35

Оценка готовности научного проекта к коммерциализации (уровень имеющихся знаний) определяется суммарным количеством баллов по каждому из направлений. Итоговый балл по каждому направлению равен 35 баллам. Если значение принадлежит диапазону от 30 до 44 баллов, то перспективность разработки считается средней. Можно сделать вывод, что для дальнейшей коммерциализации научно-исследовательского проекта необходимо вовлечение сторонних специалистов в этой области.

Такой метод коммерциализации как инжиниринг для данного научно-исследовательского проекта является наиболее подходящим. Инжиниринг предполагает предоставление заказчику на основе договора комплекса или отдельных видов инженерно-технических услуг. Данный метод дает

возможность доработать и адаптировать программный продукт к конкретной среде, что может потребоваться в процессе использования.

## **5.2 Инициация научно-исследовательского проекта**

Группа процессов инициации включает процессы, выполняемые для определения нового научно-исследовательского проекта или новой фазы существующего. В рамках процессов инициации определяются цели, а также заинтересованные стороны проекта.

### **5.2.1 Цели и результаты научно-исследовательского проекта**

В данном разделе приводится информация о заинтересованных сторонах научно-исследовательского проекта, иерархии целей проекта и критериях достижения целей. Заинтересованные стороны проекта – это лица или организации, которые активно участвуют в проекте или интересы которых могут быть затронуты как положительно, так и отрицательно в ходе исполнения или в результате завершения проекта. Информация по заинтересованным сторонам научно-исследовательского проекта представлена в таблице 7.

Таблица 7 – Заинтересованные стороны научно-исследовательского проекта

Заинтересованные стороны научно-исследовательского проекта	Ожидания заинтересованных сторон
Отделение информационных технологий ТПУ	Научные публикации Защита магистерской диссертации
Медицинский персонал	Уменьшение времени работы с медицинской документацией, благодаря автоматизации процесса степеней тяжести заболевания по данным из историй болезни пациентов с бактериологическими инфекциями
Пациенты	Получение более качественных консультаций врачей, благодаря уменьшению работы врача с документацией
Научное сообщество	Алгоритм классификации степеней тяжести заболевания пациентов с бактериологическими инфекциями

В таблице 8 представлена информация об иерархии целей научно-исследовательского проекта и критериях их достижения.

Таблица 8 – Иерархия целей научно-исследовательского проекта и критерии их достижения

Цели проекта:	Разработать алгоритмическое и программное обеспечение анализа и визуализации результатов медицинских осмотров
Ожидаемые результаты проекта:	Алгоритмическое и программное обеспечение анализа и визуализации результатов медицинских осмотров
Требования к результату проекта:	Данные подготовлены для анализа, целевая переменная выбрана
	Построен классификатор, оценивающий степень тяжести заболевания по данным пациентов из историй болезни
	Бесперебойная работа программных модулей проекта
	Формализованное описание работы программных модулей проекта

## 5.2.2 Организационная структура научно-исследовательского проекта

На данном этапе работы определяется состав рабочей группы научно-исследовательского проекта, определяются роли каждого участника в данном проекте. В таблице 9 определены участники научно-исследовательского проекта и их роли.

Таблица 9 – Рабочая группа научно-исследовательского проекта

№ п/п	ФИО, основное место работы, должность	Роль в проекте	Функции
1	Аксёнов Сергей Владимирович, ОИТ ИШИТР ТПУ, доцент	Научный руководитель	Составление научных задач, контроль выполнения проекта, проверка разработки, проверка документации
2	Кашеева Евгения Викторовна	Инженер	Проектирование, реализация

Данный раздел отражает тот факт, что выполняемая работа имеет довольно большой объём. Заинтересованные стороны научно-исследовательского проекта ожидают достаточно высококачественные результаты, которые необходимо достичь исполнителю.

## 5.3 Планирование управления научно-исследовательским проектом

### 5.3.1 План научно-исследовательского проекта

При планировании научно-исследовательского проекта были построены календарный план и диаграмма Ганта (рисунок 10). Календарный план включает информацию об этапах работы, длительности их выполнения и датах начала и завершения выполнения.

Диаграмма Ганта является наиболее удобным и наглядным способом отслеживания выполнения проектной работы [10]. Данная диаграмма представляет собой ленточный график, на котором работы представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

№	Название задачи	Начало	Окончание	Длительность	2020			
					фев	мар	апр	май
1	Составление и утверждение темы	03.02.2020	07.02.2020	5д	■ Аксёнов С.В.			
2	Разработка календарного плана	10.02.2020	14.02.2020	5д	■ Аксёнов С.В.; Кашеева Е.В.			
3	Подбор и изучение литературы	17.02.2020	06.03.2020	15д	■ Кашеева Е.В.			
4	Изучение предметной области	09.03.2020	20.03.2020	10д	■ Кашеева Е.В.			
5	Подготовка данных для анализа	23.03.2020	10.04.2020	15д	■ Кашеева Е.В.			
6	Поиск целевой переменной	13.04.2020	17.04.2020	5д	■ Кашеева Е.В.			
7	Построение классификатора	20.04.2020	30.04.2020	9д	■ Кашеева Е.В.			
8	Согласование выполненной работы с научным руководителем	04.05.2020	08.05.2020	5д	■ Аксёнов С.В.; Кашеева Е.В.			
9	Выполнение разделов «Финансовый менеджмент» и «Социальная ответственность»	11.05.2020	22.05.2020	10д	■ Кашеева Е.В.			
10	Подведение итогов, оформление работы	25.05.2020	29.05.2020	5д	■ Аксёнов С.В.; Кашеева Е.В.			

Рисунок 10 – Диаграмма Ганта

Из диаграммы Ганта наглядно видны границы этапов научно-исследовательского проекта, длительность выполнения которого составила 4 месяца. Дата начала выполнения проекта – 3 февраля, дата окончания – 29 мая 2020 года, общее количество дней – 84, суббота и воскресенье – выходные. Также на диаграмме представлена информация об исполнителях каждого отдельного этапа работы.

### 5.3.2 Бюджет научно-исследовательского проекта

В процессе планирования бюджета научно-исследовательского проекта необходимо обеспечить полное и достоверное отражение всех видов расходов, которые связаны с его выполнением. При формировании бюджета научно-исследовательского проекта используется следующая группировка затрат по статьям:

– материальные затраты научно-исследовательского проекта;

- затраты на электроэнергию, мобильную связь и Интернет;
- затраты на специальное оборудование для научных (экспериментальных) работ;
- основная заработная плата исполнителей;
- дополнительная заработная плата исполнителей;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

### 5.3.2.1 Расчет материальных затрат

К материальным затратам относятся приобретаемые сырье и материалы, канцелярские принадлежности и т.д. Материальные затраты на проведение данного научно-исследовательского проекта представлены в таблице 10.

Таблица 10 – Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед., руб.	Затраты, руб.
Ручка шариковая	шт.	2	25	50
Тетрадь в клетку, 24 листа	шт.	1	18	18
Бумага А4, 500 листов)	пачка	1	291	291
Картридж черный PG-10	шт.	1	1299	1299
Итого, руб.				1658

Материальные расходы составили 1658 рублей.

### 5.3.2.2 Расчет затрат на электроэнергию, мобильную связь и интернет

Большая часть работы была выполнена в комнате общежития Томского политехнического университета. В стоимость проживания, помимо прочего, входит плата за электроэнергию, что составляет 93,18 руб. в месяц. Так, учитывая, что работа над проектом длилась 4 месяца, затраты на электроэнергию составили 372,72 руб.

Для работы над проектом также необходим был доступ к сети Интернет, что обеспечивалось посредством раздачи через Wi-Fi с мобильного устройства. Для связи с руководителем необходима была мобильная связь. Ежесуточная

плата по тарифу МТС Smart составляет 8,4 руб. Количество дней работы над проектом, рассчитанное ранее, составило 84 дня. Общие затраты на мобильную связь и интернет составили 705,6 руб.

### 5.3.2.3 Амортизационные затраты

При выполнении работы использовался ноутбук, первоначальная стоимость которого составляет 38990 руб. Срок полезного использования составляет 3 года, компьютер используется 4 месяца согласно плану работ. Норма амортизации ( $A_n$ ), рассчитанная по следующей формуле, составила 2,8 %:

$$A_n = \frac{1}{n} \cdot 100\%, \quad (5)$$

где  $n$  – количество месяцев полезного срока эксплуатации.

Ежемесячные амортизационные отчисления ( $A_m$ ) составили 1091,72 руб., формула для расчета следующая:

$$A_m = A_n \cdot ПС, \quad (6)$$

где  $ПС$  – первоначальная стоимость основного средства.

Итоговая сумма амортизации ноутбука составила 4366,88 руб., что рассчитывается как произведение ежемесячных амортизационных отчислений и количества месяцев использования ноутбука.

### 5.3.2.4 Заработная плата исполнителей

Заработная плата рассчитывается из суммы основной и дополнительной заработной платы научного руководителя и исполнителя.

Формула для расчета основной заработной платы следующая:

$$З_{осн} = З_{дн} \cdot T_p \cdot (1 + K_{np} + K_d) \cdot K_p, \quad (7)$$

где  $З_{дн}$  – среднедневная заработная плата, руб.;  $K_{np}$  – премиальный коэффициент (30% от  $З_{дн}$ );  $K_d$  – коэффициент доплат и надбавок составляет 0,3;  $K_p$  – районный коэффициент (для Томска 1,3);  $T_p$  – продолжительность работ, выполняемых работником, раб. дни.

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{он}} = \frac{Z_m \cdot M}{F_o}, \quad (8)$$

где  $Z_m$  – оклад работника за месяц, руб.;  $M$  – количество месяцев работы без отпуска в течение года: при отпуске в 24 рабочих дня  $M=11,2$ , 5-дневная неделя;  $F_o$  – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

Баланс рабочего времени представлен в таблице 11.

Таблица 11 – Баланс рабочего времени

Показатели рабочего времени	Количество дней
Календарные дни	366
Нерабочие дни (выходные, праздничные)	118
Потери рабочего времени (отпуск, невыходы по болезни)	24
Действительный годовой фонд рабочего времени	224

Для расчета заработной платы инженера возьмем оклад, равный 21760 руб., для зарплаты руководителя – 33664 руб. В таблице 12 представлен расчет основной заработной платы.

Таблица 12 – Расчет основной заработной платы

Исполнители	$Z_{\text{он}}$ , руб.	$K_{\text{пр}}$	$K_o$	$K_p$	$T_p$	$Z_{\text{осн}}$ , руб.
Инженер	1088	0	0	1,3	84	118809,6
Руководитель	1683,2	0,3	0,3	1,3	20	63961,6

В дополнительную зарплату входят суммы выплат, предусмотренные трудовым кодексом, например, оплата ежегодных и дополнительных отпусков, оплата времени, связанного с выполнением государственных и общественных обязанностей и т.д. Запланируем дополнительную заработную плату в размере 15% от основной зарплаты исполнителей. В таблице 13 представлен расчет затрат на заработную плату исполнителей.

Таблица 13 – Расчет затрат на заработную плату исполнителей

Исполнители	$Z_{\text{осн}}$ , руб.	$Z_{\text{доп}}$ , руб.	$Z_{\text{зн}}$ , руб.
Инженер	118809,6	17821,44	136631,04
Научный руководитель	63961,6	9594,24	73555,84

### 5.3.2.5 Отчисления во внебюджетные фонды (страховые отчисления)

Размер страховых отчислений составляет 30% от суммы основной и дополнительной заработной платы исполнителя. В таблице 14 представлен расчет суммы страховых отчислений.

Таблица 14 – Расчет суммы страховых отчислений

Исполнители	$Z_{зн}$ , руб.	Страховые отчисления
Инженер	136631,04	40989,31
Научный руководитель	73555,84	22066,75
Итого	210186,88	63056,06

### 5.3.2.6 Накладные расходы

К накладным расходам относятся расходы проекта, которые не учтены в предыдущих статьях. Величина накладных расходов рассчитывается как 16 % от предыдущих расходов.

$$(1658 + 372,72 + 705,6 + 4366,88 + 210186,88 + 63056,06) \cdot 0,16 = 280\,346,14 \cdot 0,16 = 44855,38$$

### 5.3.2.7. Формирование бюджета затрат научно-исследовательского проекта

В таблице 15 представлена информация о бюджете научно-исследовательского проекта.

Таблица 15 – Бюджет научно-исследовательского проекта

Наименование	Сумма, руб.	%
Материальные затраты	1658	0,51
Затраты на электроэнергию, мобильную связь и интернет	1078,32	0,33
Затраты на специальное оборудование	4366,88	1,34
Затраты на основную заработную плату	182771,2	56,2
Затраты на дополнительную заработную плату	27415,68	8,43
Страховые взносы	63056,06	19,39
Накладные расходы	44855,38	13,8
Общий бюджет	325201,52	100

Таким образом, большая часть расходов по проекту относится к категории затрат на основную заработную плату исполнителей, что составляет 56,2 %.

### 5.3.3 Риски научно-исследовательского проекта

При разработке научно-исследовательского проекта следует понимать и учитывать возможные риски. Риски в реализации проекта включают в себя возможные неопределенные события, которые могут возникнуть в проекте и вызвать последствия, которые повлекут за собой нежелательные эффекты. В таблице 16 представлены результаты оценки рисков проекта. Для каждого из них даны рекомендации по смягчению их воздействия.

Таблица 16 – Реестр рисков

№	Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска	Способы смягчения риска	Условия наступления
1	Кад- ровый риск	Отсутствие заинтересованных исполнителей проекта	3	5	Существенный риск	Повышение мотивации исполнителей проекта	Потеря интереса исполнителей к деятельности проекта
2	Тех- ниче- ский риск	Потеря файлов проекта	2	5	Существенный риск	Регулярное создание резервных копий файлов проекта	Отказ используемого оборудования
3	Дос- туп к дан- ным	Отсутствие данных для работы системы	2	5	Существенный риск	Заключение официального договора на доступ к данным	Отсутствие доступа к данным истории

Из анализа реестра рисков можно заключить, что первым и вторым типами рисков обладает практически каждый проект. Риск же потери доступа к данным во время выполнения данного научно-исследовательского проекта существенен для реализации, однако маловероятен.

### 5.3.4 Описание потенциального эффекта

В результате проделанной в рамках раздела работы, можно сделать выводы о том, что на данный момент невозможно оценить экономический эффект разработки до её внедрения. Однако согласно оценке научно-исследовательского проекта по технологии «QuaD» разработка считается перспективной. Для коммерциализации научно-исследовательского проекта

необходимо вовлечение сторонних специалистов в этой области. Основные работы в рамках данного научно-исследовательского проекта проводились в период с 3 февраля по 29 мая 2020 года. Команда проекта состоит из инженера и научного руководителя. Общий бюджет научно-исследовательского проекта составил 325201,52 руб. Основная часть затрат приходится на заработную плату исполнителей проекта. Все рассмотренные риски научно-исследовательского проекта являются существенными для его реализации, однако вероятность их наступления достаточно мала.

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

<b>Группа</b>	<b>ФИО</b>
8ПМ8И	Кащева Евгения Викторовна

<b>Инженерная школа</b>	<b>ИШИТР</b>	<b>Отделение (НОЦ)</b>	Информационных технологий
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.09.04 Программная инженерия

Тема ВКР:

Алгоритмическое и программное обеспечение анализа и визуализации результатов медицинских осмотров	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Алгоритм классификации степеней тяжести заболевания пациентов с бактериологическими инфекциями. Алгоритм может применяться в системах принятия решений врача.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	<ul style="list-style-type: none"> <li>– Трудовой Кодекс Российской Федерации</li> <li>– Федеральный закон «О специальной оценке условий труда»</li> <li>– СанПиН 2.2.4.548-96 «Гигиенические требования к микроклимату производственных помещений»</li> <li>– СанПиН 2.2.4.3359-16 «Санитарно-эпидемиологические требования к физическим факторам на рабочих местах»</li> <li>– СП 52.13330.2016 «Естественное и искусственное освещение». Актуализированная редакция СНиП 23-05-95</li> <li>– СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» (с изменениями на 21 июня 2016 года)</li> </ul>
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Вредные факторы: <ul style="list-style-type: none"> <li>– Отклонение показателей микроклимата</li> <li>– Превышение уровня шума</li> <li>– Отсутствие или недостаток естественного света</li> <li>– Повышенная напряженность магнитного поля</li> <li>– Психофизиологический фактор</li> </ul> Опасный фактор: <ul style="list-style-type: none"> <li>– Электрический ток</li> </ul>
<b>3. Экологическая безопасность:</b>	Анализ негативного воздействия на окружающую природную среду: утилизация люминесцентных ламп, компьютеров и другой оргтехники
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Удар молнии, взрыв, наводнение. Наиболее типичная ЧС – пожар.

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Горбенко Михаил Владимирович	к.т.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ПМ8И	Кащева Евгения Викторовна		

## **6 Социальная ответственность**

### **Введение**

Данная работа посвящена разработке алгоритма классификации степеней тяжести заболевания по данным из историй болезни пациентов с бактериологическими инфекциями. Данный алгоритм может найти свое применение в системах поддержки принятия решений врача, которые смогут помочь специалисту сократить время, а также увеличить точность принятия решений о способе лечения пациентов на основе схожих случаев.

Работа выполнялась с использованием ЭВМ на базе Отделения информационных технологий Инженерной школы информационных технологий и робототехники Томского политехнического университета (ТПУ), данные для работы были предоставлены Отделением инфекционных заболеваний Сибирского государственного медицинского университета.

В разделе будут рассмотрены опасные и вредные факторы, оказывающие влияние на производственную деятельность инженера-программиста. Исследовано рабочее место программиста и помещение, в котором оно находится. Разработка осуществлялась в компьютерном классе Кибернетического центра ТПУ, а затем в комнате общежития ТПУ. Рассмотрены воздействия объекта исследования на окружающую среду, правовые и организационные вопросы, а также мероприятия в чрезвычайных ситуациях.

### **6.1 Правовые и организационные вопросы обеспечения безопасности**

#### **6.1.1 Специальные правовые нормы трудового законодательства**

Основным документом, регулирующим отношения между работником и работодателем, является Трудовой Кодекс Российской Федерации (ТК РФ) [11]. Согласно данному документу продолжительность рабочего времени не может превышать 40 часов в неделю. Федеральный орган исполнительной власти, осуществляющий функции по выработке государственной политики и нормативно-правовому регулированию в сфере труда, определяет порядок

исчисления нормы рабочего времени в зависимости от установленной продолжительности рабочего времени в неделю.

Продолжительность рабочего дня или смены, предшествующих нерабочему праздничному дню, сокращается на один час.

Также работнику в течение рабочего дня (смены) должен предоставляться перерыв на отдых и питание продолжительностью от 30 минут до 2 часов, который не включается в рабочее время. Если продолжительность ежедневной работы не превышает 4 часов, то указанный перерыв может не предоставляться. Обязательным является предоставление всем работникам выходных дней (еженедельный непрерывный отдых).

Труд работников должен быть оплачен организацией-работодателем. Удержание заработной платы возможно только в случаях, рассмотренных в статье 137 ТК РФ.

Обработка персональных данных работника может осуществляться исключительно для обеспечения соблюдения законов и иных нормативных правовых актов, содействия работникам в трудоустройстве, получении образования и продвижении по службе, обеспечения личной безопасности работников, контроля количества и качества выполняемой работы и обеспечения сохранности имущества.

Персональные данные работника следует получать у него самого. Чтобы получить персональные данные работника от третьих лиц, необходимо уведомить об этом сотрудника, а также получить от него письменное согласие. Работодатель должен сообщить работнику о целях, предполагаемых источниках и способах получения персональных данных, а также характере подлежащих получению персональных данных и последствиях отказа работника дать письменное согласие на их получение.

Работодатель обязан обеспечить нормальные условия для выполнения работниками норм выработки. К таким условиям, в частности, относят следующие:

– исправное состояние помещений, сооружений, машин, технологической оснастки и оборудования;

– своевременное обеспечение технической и иной необходимой для работы документацией;

– условия труда, соответствующие требованиям охраны труда и безопасности производства;

– обучение безопасным методам и приемам выполнения работ и оказанию первой помощи пострадавшим на производстве, проведение инструктажа по охране труда, стажировки на рабочем месте и проверки знания требований охраны труда;

– недопущение к работе лиц, не прошедших в установленном порядке обучение и инструктаж по охране труда, стажировку и проверку знаний требований охраны труда;

– организацию контроля за состоянием условий труда на рабочих местах, а также за правильностью применения работниками средств индивидуальной и коллективной защиты;

– информирование работников об условиях и охране труда на рабочих местах, о риске повреждения здоровья, предоставляемых им гарантиях, полагающихся им компенсациях и средствах индивидуальной защиты;

– принятие мер по предотвращению аварийных ситуаций, сохранению жизни и здоровья работников при возникновении таких ситуаций, в том числе по оказанию пострадавшим первой помощи;

– санитарно-бытовое обслуживание и медицинское обеспечение работников в соответствии с требованиями охраны труда, а также доставку работников, заболевших на рабочем месте, в медицинскую организацию в случае необходимости оказания им неотложной медицинской помощи.

Федеральный закон «О специальной оценке условий труда» [12] регламентирует проведение специальной оценки в случае, если деятельность работников предприятия предусматривает непрерывную работу за компьютерными системами. По проведенной оценке устанавливаются гарантии

и компенсации работникам согласно ТК РФ. Если условия труда на рабочих местах признаны вредными, сотрудники в зависимости от степени вредности имеют право на сокращенную рабочую неделю не более 36 часов, дополнительный отпуск не менее 7 календарных дней и/или компенсацию в размере 4 % от оклада.

### 6.1.2 Эргономические требования к рабочему месту оператора ПЭВМ

Главными элементами рабочего места программиста являются стол, кресло, дисплей, клавиатура и мышь. Основным рабочим положением является положение сидя. Нормативными положениями СанПиНа [13] предъявляются определенные требования к оснащению рабочего места, предусматривающего длительную работу за персональным компьютером (Таблица 17).

Таблица 17 – Нормы оборудования рабочих мест с ПЭВМ

Параметр		Значение
Высота перегородок, разделяющих рабочие места		Не менее 1,5 метров
Стол	Ширина рабочей поверхности	От 80 до 140 см
	Глубина рабочей поверхности	От 80 до 100 см
	Высота рабочей поверхности	72,5 см
Расстояние от глаз до монитора		От 60 до 70 см
Расстояние клавиатуры от края стола		От 10 до 30 см
Стул	Ширина поверхности	От 40 см
	Глубина поверхности	От 40 см
	Регулировка высоты поверхности	От 40 до 50 см
	Угол наклона вперед	До 15 градусов
	Угол наклона назад	До 5 градусов
	Высота опорной поверхности спинки	30 плюс/минус 2 см
	Ширина опорной поверхности спинки	От 38 см
	Радиус кривизны горизонтальной плоскости спинки	40 см
	Угол наклона спинки в вертикальной плоскости	Плюс/минус 30 градусов
	Регулировка расстояния спинки от переднего края сидения	От 26 до 40 см
	Длина подлокотников	От 25 см
	Ширина подлокотников	От 5 до 7 см
	Регулировка подлокотников по высоте над сиденьем	23 плюс/минус 3 см
	Внутреннее расстояние между подлокотниками	От 35 до 50 см
Подставка для ног	Ширина	От 30 см
	Глубина	От 40 см
	Регулировка по высоте	До 15 см

Продолжение таблицы 17

	Угол наклона опорной поверхности	До 20 градусов
	Высота бортика по переднему краю	1 см

Работающий на ПЭВМ должен сидеть прямо, опираясь в области нижнего края лопаток на спинку кресла, не сутулясь, с небольшим наклоном головы вперед (до 5 – 7 градусов). Предплечья должны опираться на поверхность стола, снимая статическое напряжение плечевого пояса и рук.

Рабочие места в компьютерном классе Кибернетического центра (КЦ) ТПУ и общежития ТПУ отвечают данным требованиям.

## 6.2 Производственная безопасность

### 6.2.1 Анализ вредных и опасных факторов, которые может создать объект исследования

При разработке проекта возможно возникновение вредных и опасных факторов [14], представленных в таблице 18.

Таблица 18 – Вредные и опасные производственные факторы

Факторы	Этапы работ		Нормативные документы
	Разработка	Эксплуатация	
Отклонение показателей микроклимата	+	+	СанПиН 2.2.4.548-96 «Гигиенические требования к микроклимату производственных помещений» [15].
Превышение уровня шума	+	+	СанПиН 2.2.4.3359-16 «Санитарно-эпидемиологические требования к физическим факторам на рабочих местах» [16].
Отсутствие или недостаток естественного света	+	+	СП 52.13330.2016 «Естественное и искусственное освещение». Актуализированная редакция СНиП 23-05-95 [17].
Повышенная напряженность магнитного поля	+	+	СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» (с изменениями на 21 июня 2016 года) [13].

## Продолжение таблицы 18

Факторы	Этапы работ		Нормативные документы
	Разработка	Эксплуатация	
Психофизиологический фактор	+	+	СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» (с изменениями на 21 июня 2016 года) [13].
Электрический ток	+	+	Защитное заземление, зануление, ГОСТ 12.1.030–81 ССБТ [24]. ГОСТ Р 12.1.019-2009 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты [25].

Данные факторы могут влиять на состояние здоровья, привести к травмам или аварийной ситуации, поэтому необходимо устанавливать эффективный контроль за соблюдением норм и требований, предъявляемых к их параметрам.

### **6.2.2 Обоснование мероприятий по снижению воздействия вредных и опасных факторов**

#### **6.2.2.1 Микроклимат**

Длительное воздействие на человека неблагоприятных метеорологических условий резко ухудшает его самочувствие, снижает производительность труда и часто приводит к заболеваниям.

Значительное влияние на терморегуляцию организма оказывает влажность воздуха. Высокая относительная влажность воздуха при его высокой температуре способствует перегреванию организма. Низкая влажность вызывает пересыхание слизистых оболочек дыхательных путей. Подвижность воздуха достаточно эффективно способствует теплоотдаче, что является положительным явлением при высокой температуре окружающей среды и отрицательным – при низкой.

Для категории тяжести работ 1а (работы с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт), производимые сидя и сопровождающиеся незначительным физическим напряжением) оптимальные и допустимые

показатели температуры, относительной влажности и скорости движения воздуха в рабочей зоне производственных помещений должны соответствовать значениям, представленным в таблице 19 [15].

Таблица 19 – Оптимальные нормы микроклимата

Период года	Температура, °С	Температура поверхностей, °С	Относительная влажность, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	40-60	0,1
Теплый	23-25	22-26	40-60	0,1

Допустимые микроклиматические условия установлены по критериям допустимого теплового и функционального состояния человека на период 8-часовой рабочей смены. Они устанавливаются в случаях, когда по технологическим требованиям, технически и экономически обоснованным причинам не могут быть обеспечены оптимальные величины. В таблице 20 представлены допустимые величины показателей микроклимата на рабочих местах.

Таблица 20 – Допустимые величины показателей микроклимата

Период года	Температура воздуха, °С		Относительная влажность воздуха, %	Скорость движения воздуха, не более, м/с	
	Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин		при температуре воздуха ниже оптимальной	при температуре воздуха выше оптимальной
Холодный	20,0 – 21,9	24,1 – 25,0	15 – 75	0,1	0,1
Теплый	21,0 – 22,9	25,1 – 28,0	15 - 75	0,1	0,2

При обеспечении допустимых величин микроклимата на рабочих местах перепад температуры воздуха по высоте должен быть не более 3°С, перепад температуры воздуха по горизонтали, а также ее изменения в течение смены не должны превышать – 4°С. При этом абсолютные значения температуры воздуха не должны выходить за пределы оптимальных величин.

При температуре воздуха на рабочих местах 25°С и выше максимально допустимые значения относительной влажности воздуха не должны выходить за следующие пределы:

- 70% – при температуре воздуха 25°С,
- 65% – при температуре воздуха 26°С,

– 60% – при температуре воздуха 27°C,

– 55% – при температуре воздуха 28°C.

Для обеспечения установленных норм микроклиматических параметров и чистоты воздуха на рабочих местах и в помещениях применяется вентиляция, периодически ведется контроль влажности воздуха. В летнее время при высокой уличной температуре используется система кондиционирования. В холодное время года предусматривается система отопления. Для отопления помещений используются водяные системы центрального отопления.

Помещение, где выполнялась работа, было обследовано на соответствие данным требованиям. Результаты обследования приведены в таблице 21.

Таблица 21 – Результаты измерений параметров микроклимата

Рабочее место	Период года	Температура, °С	Относительная влажность воздуха, %	Скорость движения воздуха, не более, м/с
Аудитория КЦ ТПУ	Холодный (январь)	22,4	60	0
	Тёплый (июль)	23,0	59	0
Комната общежития ТПУ	Холодный (январь)	23,5	55	0
	Тёплый (июль)	24,7	53	0

Результаты измерений соответствуют допустимым значениям нормативов, следовательно, микроклимат помещений удовлетворяет требованиям санитарных норм и правил.

#### 6.2.2.2 Уровень шума

Раздражающее влияние на работника оказывает шум, что является причиной повышения утомляемости, а при выполнении задач, требующих сосредоточенности и внимания, может привести к росту ошибок и увеличению длительности выполнения задания. Нормативным эквивалентным уровнем звука на рабочих местах является 50 дБА [16].

Источником шумовых помех на рабочем месте могут выступать кондиционеры, вентиляционные установки, ЭВМ и ее периферийные устройства, а также серверные комнаты. Что касается комнаты общежития, то источниками шума являются холодильник, часы и ноутбук. С помощью

мобильного приложения «Sound Meter» были проведены измерения уровня шума в помещениях. Результаты измерений представлены в таблице 22.

Таблица 22 – Результаты измерений уровня звука

Рабочее место	Рабочая операция	Уровень звука, дБА		Продолжительность операции, мин	
		Результаты измерений	Эквивалентный уровень за операцию	Результаты наблюдений	Средняя
Аудитория КЦ ТПУ	Работа за персональным компьютером	43, 48, 45	45,3	20, 20, 20	20
Комната общежития ТПУ	Работа за персональным компьютером	30, 45, 37	37,3	20, 20, 20	20

В результате вычисления на основе измеренных величин показателей шума на рабочих местах эквивалентные уровни звука за 8-часовой рабочий день составляют 45,3 и 37,3 дБА, что соответствует нормативным значениям.

Снизить уровень шума в помещении можно с помощью звукопоглощающих материалов с максимальными коэффициентами звукопоглощения в области частот 63 – 8000 Гц для отделки стен и потолка помещений. Также дополнительный звукопоглощающий эффект создается однотонными занавесками из плотной ткани, повешенными в складку на расстоянии 15 – 20 см от ограждения. Ширина занавески должна быть в 2 раза больше ширины окна.

### 6.2.2.3 Отсутствие или недостаток естественного света

В помещении при работе с ПК должно быть естественное и искусственное освещение. Естественное освещение обеспечивается через оконные проемы с коэффициентом естественного освещения не ниже 1,2% в зонах с устойчивым снежным покровом и не ниже 1,5% на остальной территории. Световой поток из оконного проема должен падать на рабочее место оператора с левой стороны.

Для определения приемлемого уровня освещенности в помещении необходимо следующее: определить требуемый уровень освещенности внешними источниками света; если требуемый уровень освещенности не

приемлем, необходимо найти способ сохранения требуемого контраста изображения другими средствами.

Рекомендуемые соотношения яркостей в поле зрения между рабочими поверхностями не должно превышать 1:3 – 1:5, между рабочими поверхностями и поверхностями стен и оборудования – 1:10.

Освещенность на поверхности стола в зоне размещения документа должна быть 300 – 500 лк [13]. Допускается установка светильников местного освещения для подсветки документов. Местное освещение не должно создавать бликов на поверхности экрана и увеличивать освещенность экрана более 300 лк. Прямую блеклость от источников освещения следует ограничить. Яркость светящихся поверхностей (окна, светильники), находящихся в поле зрения, должна быть не более 200 кд/м<sup>2</sup>.

В таблице 23 представлены нормы на освещение для операторов поста управления [17].

Таблица 23 – Нормы на освещение для оператора

Характер зрительной работы	Разряд зрительной работы	Подразряд зрительной работы	Искусственное освещение		Естественное освещение, % при боковом направлении
			Освещенность при системе общего освещения, лк	Коэффициент пульсации, %	
Различие объектов высокой точности	Б	1	300	15	1

Ниже приведем расчет системы искусственного освещения для компьютерного класса КЦ ТПУ и комнаты общежития ТПУ методом светового потока, предназначенного для расчета освещенности общего равномерного освещения горизонтальных поверхностей. Помещения, для которых будут производиться расчеты, прямоугольные со следующими параметрами: аудитория КЦ ТПУ имеет длину 5 м, ширину – 7 м, высоту – 4 м, количество ламп – 12 шт.; комната общежития ТПУ имеет длину 3 м, ширину – 6 м, высоту – 2,5 м, количество ламп – 8 шт. Уровень рабочей поверхности над полом должен составлять 0,8 м, установленная минимальная норма освещенности – 300 Лк.

Формула для определения светового потока лампы накаливания или группы люминесцентных ламп светильника следующая [26]:

$$\Phi = \frac{E_n \cdot S \cdot K_z \cdot Z \cdot 100}{n \cdot \eta}, \quad (9)$$

где  $E_n$  – нормируемая минимальная освещенность [17];  $S$  – площадь освещаемого помещения;  $K_z$  – коэффициент запаса, учитывающий загрязнение светильника, пыли (для помещений с малым выделением пыли равен 1,5);  $Z$  – коэффициент неравномерного освещения (для люминесцентных ламп равен 1,1);  $n$  – количество светильников;  $\eta$  – коэффициент использования светового потока, %.

Коэффициент использования светового потока показывает, какая часть светового потока ламп попадает на рабочую поверхность. Он зависит от индекса помещения  $i$ , типа светильника, высоты светильников над рабочей поверхностью  $h$  и коэффициентов отражения стен, потолка и рабочей поверхности. В Таблице 24 представлены значения коэффициентов использования светового потока [26].

Таблица 24 – Значения коэффициентов использования светового потока

	потолок	0,8	0,7	0,7	0,5	0,5	0,5
	стены	0,5	0,5	0,3	0,5	0,3	0,3
	рабочая поверхность	0,3	0,3	0,3	0,3	0,3	0,1
Индекс помещения	0,6	0,33	0,32	0,25	0,3	0,24	0,24
	0,8	0,41	0,39	0,32	0,36	0,3	0,29
	1	0,47	0,45	0,38	0,42	0,35	0,34
	1,25	0,53	0,51	0,44	0,47	0,41	0,39
	1,5	0,58	0,55	0,48	0,51	0,45	0,43
	2	0,65	0,62	0,56	0,57	0,52	0,49
	2,5	0,7	0,67	0,61	0,61	0,56	0,53
	3	0,64	0,71	0,65	0,64	0,6	0,56
	4	0,79	0,75	0,7	0,68	0,64	0,6
5	0,83	0,78	0,74	0,71	0,68	0,62	

Формула для определения индекса помещения следующая [26]:

$$i = \frac{S}{h \cdot (A + B)}, \quad (10)$$

где  $h$  – расчетная высота подвеса светильников над рабочей поверхностью;  $A$  – длина помещения;  $B$  – ширина помещения.

Произведем расчеты высоты подвеса светильников над рабочей поверхностью для аудитории и комнаты:

$$h_a = H - 0,8 = 4 - 0,8 = 3,2$$

$$h_k = H - 0,8 = 2,5 - 0,8 = 1,7$$

Рассчитаем индексы помещений:

$$i_a = \frac{5 \cdot 7}{3,2 \cdot (5 + 7)} = \frac{35}{3,2 \cdot 12} = 0,91$$

$$i_k = \frac{3 \cdot 6}{1,7 \cdot (3 + 6)} = \frac{18}{1,7 \cdot 9} = 1,18$$

Найдем коэффициенты отражения поверхностей стен, потолка и рабочей поверхности, в таблице 25 представлены коэффициенты отражения для различных поверхностей [26].

Таблица 25 – Коэффициенты отражения

Характеристика поверхности	Коэффициент отражения, %
Белая поверхность (побелка, белые обои, белая плитка)	70
Светлая поверхность (светлые обои, светлый потолок, светлая краска)	50
Поверхность средней светлости (светлый паркет, светлый линолеум, несветлая краска, несветлые обои)	30
Темная поверхность (темный паркет, темная краска, темный линолеум, темные обои)	10

Так как поверхности стен аудитории и комнаты окрашены в светлый цвет, то коэффициенты отражения стен равны 50%, поверхности потолков аудитории и комнаты также светлого цвета, коэффициенты отражения потолков равны 50%, рабочие поверхности средней светлости, коэффициенты отражения рабочих поверхностей равны 30%. Учитывая коэффициенты отражения поверхностей стен, потолка, рабочей поверхности и индекс помещения  $i$ , определяем значения коэффициентов  $\eta_a = 36\%$ ,  $\eta_k = 42\%$ .

Подставив значения в формулу (9), рассчитываем световой поток одного источника света:

$$\Phi_a = \frac{300 \cdot 35 \cdot 1,5 \cdot 1,1}{12 \cdot 0,36} = 4010$$

$$\Phi_k = \frac{300 \cdot 18 \cdot 1,5 \cdot 1,1}{8 \cdot 0,42} = 2651$$

По полученному световому потоку подбираем лампы, наиболее подходящими являются светодиодная лампа Smartbuy LED HP50W (4000 лм) и светодиодная лампа КОСМОС LED 30Вт E27 (2650 лм).

Выразим  $E$ :

$$E = \frac{(F \cdot N \cdot \eta)}{(k)}, \quad (11)$$

Рассчитаем норму освещенности:

$$E_a = \frac{(4000 \cdot 12 \cdot 0,36)}{(1,5 \cdot 35 \cdot 1,1)} = 299,2$$

$$E_k = \frac{(2650 \cdot 8 \cdot 0,42)}{(1,5 \cdot 18 \cdot 1,1)} = 229,8$$

Как видно из расчета, минимальная освещенность в пределах нормы [17].

Для того чтобы доказать, что использование люминесцентных ламп Smartbuy и КОСМОС является наиболее рациональным, рассчитаем необходимое количество светильников по формуле:

$$N = \frac{(E \cdot k \cdot S \cdot Z)}{(n \cdot \eta \cdot F)}, \quad (12)$$

где  $E$  – норма освещенности, 300 Лк;  $k$  – коэффициент запаса, учитывающий старение ламп и загрязнение светильников, 1,5;  $S$  – площадь помещения;  $Z$  – коэффициент неравномерности освещения, 1,1;  $n$  – число рядов светильников, 2;  $\eta$  – коэффициент использования светового потока;  $F$  – световой поток, излучаемый светильником.

Рассчитаем количество ламп:

$$N_a = \frac{300 \cdot 1,5 \cdot 35 \cdot 1,1}{0,36 \cdot 4000} = 12$$

$$N_k = \frac{300 \cdot 1,5 \cdot 18 \cdot 1,1}{0,42 \cdot 2650} = 8$$

В аудитории находятся три светильника, в каждом из которых по 4 лампы, в комнате общежития находятся два светильника, также содержащие по 4 лампы. Расчетное количество ламп соответствует их реальному количеству

при норме освещенности 300 Лк, что говорит о соблюдении норм по искусственному освещению [17].

#### **6.2.2.4 Повышенная напряженность магнитного поля**

Электромагнитные поля, характеризующиеся напряженностями электрических и магнитных полей, наиболее вредны для организма человека. Основным источником этих проблем, связанных с охраной здоровья людей, использующих в своей работе автоматизированные информационные системы на основе персональных компьютеров, являются дисплеи (мониторы), они представляют собой источники наиболее вредных излучений, неблагоприятно влияющих на здоровье человека. Предельно допустимые значения излучений от ЭВМ [13] приведены в таблице 26.

Таблица 26 – Допустимые уровни электромагнитного поля

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	В диапазоне частот 5 Гц – 2 кГц	25 В/м
	В диапазоне частот 2 кГц – 400 кГц	2,5 В/м
Плотность магнитного потока	В диапазоне частот 5 Гц – 2 кГц	250 нТл
	В диапазоне частот 2 кГц – 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500 В

Для оценки соблюдения уровней электромагнитного поля необходим производственный контроль. В случае их превышения необходимо проводить организационно-технические мероприятия (защита временем, расстоянием, экранирование источника, замена оборудования, использование средств индивидуальной защиты).

#### **6.2.2.5 Психофизиологический фактор**

Настоящей работе сопутствует ряд вредных психофизиологических факторов: напряжение зрения и внимания; эмоциональные, интеллектуальные и длительные статические нагрузки; большая монотонность труда; большой объем обрабатываемой информации в единицу времени; нерациональная организация рабочего места.

В результате воздействия данных факторов к концу рабочего дня рабочий испытывает неприятные ощущения: переутомление глаз, головную

боль, тянущие боли в мышцах шеи, рук и спины, снижение концентрации внимания.

В целом продолжительность непрерывной работы за компьютером не должна превышать 2-х часов. Основная работа за компьютером предусматривает не менее 50% времени в течение рабочей смены или рабочего дня нахождения за ним. Время перерыва зависит от вида и сложности осуществляемой работы путем деления на группы. Выделяют 3 группы: А (работа по считыванию информации с экрана компьютера с предварительным запросом), Б (работа по вводу информации), В (творческая работа в режиме диалога с компьютером). Во время перерывов следует выполнять специальную гимнастику для снятия напряжения с глаз [13]. Выполнять какую-либо работу, не связанную с компьютером, во время перерыва нельзя. Потому как перерыв приравнивается к времени отдыха. А время отдыха – это свободное от исполнения трудовых обязанностей время, которое работник может использовать по своему усмотрению [11].

#### **6.2.2.6 Электрический ток**

Степень опасных воздействий на человека электрического тока зависит от следующих факторов:

- род и величина напряжения и тока;
- частота электрического тока;
- путь прохождения тока через тело человека;
- продолжительность воздействия на организм человека;
- условия внешней среды.

Согласно правилам устройства электроустановок аудиторию КЦ ТПУ и комнату общежития ТПУ по степени опасности поражения электрическим током можно отнести к классу помещений без повышенной опасности.

Основные мероприятия по защите от электропоражения следующие:

- обеспечение недоступности токоведущих частей путем использования изоляции в корпусах оборудования;

- применение средств коллективной защиты от поражения электрическим током;
- защитное заземление, зануление [24];
- защитное отключение;
- использование устройств бесперебойного питания.

Технические способы и средства применяют отдельно или в сочетании друг с другом так, чтобы обеспечивалась оптимальная защита.

Электробезопасность должна обеспечиваться [25]:

- конструкцией электроустановок;
- техническими способами и средствами защиты;
- организационными и техническими мероприятиями.

### **6.3 Экологическая безопасность**

Деятельность по разработке ПО не связана с производством, поэтому влияние на окружающую среду минимально. При работе над проектом применяются рекомендации по минимизации влияния на окружающую среду.

С учетом того, что наиболее значительную часть твердых бытовых отходов (до 40 % в развитых странах) составляет бумага и картон — бумага для печати, упаковка и упаковочные материалы, актуальным является осуществление их утилизации для повторного использования в промышленном производстве, что является наиболее экономически эффективным способом обращения с отходами [18].

При завершении срока службы ПК их можно отнести к отходам электронной промышленности. Переработка таких отходов осуществляется разделением на однородные компоненты, химическим выделением пригодных для дальнейшего использования компонентов и направлением их для дальнейшего использования [19].

Перечень элементов и содержащее их отработанное электротехническое и электронное оборудование, которые должны быть отдельно собраны при выводе отработавшего электротехнического и электронного оборудования из эксплуатации:

- конденсаторы, содержащие ПХБ;
- печатные платы и других устройств с площадью поверхности больше 10 см<sup>2</sup>;
- картриджи;
- пластик;
- электронно-лучевые трубки;
- элементы отработавшего электротехнического и электронного оборудования;
- газоразрядные лампы;
- жидкокристаллические экраны (если необходимо, вместе с корпусом) с поверхностью более 100 см<sup>2</sup> и все экраны с подсветкой газоразрядными лампами;
- внешние электрические кабели;
- элементы, содержащие огнеупорные керамические слои;
- конденсаторы, содержащие электролит (размер хотя бы одной из сторон конденсатора должен быть 25 мм или более).

Люминесцентные лампы относят к ртутьсодержащим отходам, и для их утилизации действует Постановление Правительства РФ [20]. Устанавливается порядок обращения с отходами производства и потребления в части осветительных устройств, электрических ламп, ненадлежащие сбор, накопление, использование, обезвреживание, транспортирование и размещение которых может повлечь причинение вреда жизни, здоровью граждан, вреда животным, растениям и окружающей среде.

Не допускается самостоятельное обезвреживание, использование, транспортирование и размещение отработанных ртутьсодержащих ламп потребителями отработанных ртутьсодержащих ламп, а также их накопление в местах, являющихся общим имуществом собственников помещений многоквартирного дома, за исключением размещения в местах первичного сбора и размещения и транспортирования до них.

Сбор отработанных ртутьсодержащих ламп у потребителей осуществляют специализированные организации. Отходы, не подлежащие переработке и вторичному использованию подлежат захоронению на полигонах.

## **6.4 Безопасность в чрезвычайных ситуациях**

### **6.4.1 Анализ вероятных ЧС, которые может инициировать объект исследований**

Перечень возможных ЧС на объекте исследования может быть достаточно широк. Ограничиваясь местоположением объекта и условиями его эксплуатации, его можно представить следующим (ориентировочным) вариантом:

- наводнение;
- удар молнии;
- пожар на объекте;
- взрыв.

Аудитория КЦ ТПУ, относится к категории ВЗ по пожароопасности, содержит вещества и материалы, способные гореть при взаимодействии с водой, кислородом воздуха или друг с другом [21]. Комната общежития относится к жилым помещениям, которые не подвергаются категорированию по пожаровзрывоопасности.

Для минимизации возможности возникновения пожара необходимо проводить пожарную профилактику. Под пожарной профилактикой понимают комплекс организационных и технических мероприятий, направленных на обеспечение безопасности людей, на предотвращении пожара, ограничение его распространения, а также создание условий для успешного тушения пожара. Для профилактики пожара чрезвычайно важна правильная оценка пожароопасности, определение опасных факторов и обоснование способов и средств предупреждения пожара и защиты от него.

Одно из условий обеспечения пожаробезопасности - ликвидация возможных источников воспламенения. Обогревание помещения открытыми

электронагревательными приборами могут привести к пожару, т.к. в помещении находятся бумажные документы и справочная литература. Следовательно, использование открытого нагревательного прибора неприемлемо.

#### **6.4.2 Меры по предотвращению и ликвидации ЧС и их последствий**

В целях предотвращения пожара предлагаются следующие мероприятия:

- проведение с сотрудниками противопожарного инструктажа;
- проведение планового осмотра и своевременное устранение всех неисправностей в электроприборах;
- предотвращение небезопасного хранения легковоспламеняющихся жидкостей;
- оснащение помещения автоматической системой обнаружения пожара;
- оснащение помещения автоматической системой оповещения о пожаре.

Помещения с категорией В3 должны оснащаться системой автоматической пожарной сигнализации и система оповещения и управления эвакуацией людей при пожарах [22, 23].

Аудитория КЦ ТПУ оснащена одним пожарным, а также двумя дымовыми оповещателями, для информирования о пожаре установлена звуковая сирена. Комната общежития оснащена одним дымовым датчиком, а также в секции находится одна звуковая сирена.

#### **6.5 Выводы**

В результате проведенного анализа были выявлены вредные и опасные производственные факторы для аудитории КЦ ТПУ, комнаты общежития и работы, связанной с проектом. Были определены значения нормативных показателей для вредных и опасных факторов, которые сравнивались со значениями, действующими при разработке проекта. Приведены рекомендации по улучшению условий труда и минимизации влияния вредных и опасных факторов на работника. Проведен анализ воздействия на окружающую среду и

обозначены проблемы утилизации отходов. Определена категория помещения по пожароопасности, а также действия по минимизации риска возникновения пожара. Также были рассмотрены правовые нормы трудового законодательства, применимые к условиям настоящего проекта, определены основные требования к организации рабочих мест. После проведенного анализа можно сделать вывод, что рабочие места соответствуют нормативным требованиям производственной безопасности и охраны труда.

## **Заключение**

На протяжении выполнения данной работы поставленные задачи были выполнены. А именно, была проведена подготовка текстовых данных из медицинских документов к анализу, а также построен классификатор степеней заболевания пациентов, страдающих рожистым воспалением.

Предварительно загруженные данные были проверены на наличие пропущенных значений и выбросов. Затем текстовые данные подверглись токенизации, лемматизации, удалению стоп-слов и транслитерации, после чего текст был переформатирован в числовые значения, полученные в результате расчета весов слов методом «TF-IDF». Данная метрика позволила учесть частоту наиболее значимых слов, уменьшая вес слов, встречающихся в каждом документе.

Набор данных был разделен на тренировочное и тестовое подмножества. Было построено несколько моделей для определения степеней тяжести рожистого воспаления. Из полученных моделей была выбрана лучшая – градиентный бустинг. Для выбранной модели был проведен подбор оптимальных параметров. После чего была проведена оценка качества работы модели классификации с помощью матрицы ошибок и метрик качества, так верность работы модели составила 95%, полнота – 96%, точность – 94% и метрика F1 – 95%, что является хорошими результатами.

В результате проделанной работы для классификации степеней тяжести заболеваний пациентов, страдающих рожистым воспалением, предлагается использование классификатора «XGBoostClassifier». Использование разработанной модели поможет в построении эффективных медицинских информационных систем.

Также разработаны следующие разделы: «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение», «Социальная ответственность», а также раздел на иностранном языке (английский) – «Big Data Analytics in Medicine», приведенный в Приложении А.

## **Список научных достижений и публикаций**

1. Диплом II степени за лучший доклад на Международной научно-технической конференции студентов, аспирантов и молодых ученых "Научная сессия ТУСУР", 22-24 мая 2019 г.

2. Диплом III степени за доклад на Международной научной конференции "Информационные технологии в науке, управлении, социальной сфере и медицине", 14-19 октября 2019 г.

3. Семёнов С. А. , Кашеева Е. В. , Аксёнов С. В. Алгоритмическое и программное обеспечение сопровождения анализа результатов радиологических исследований // Молодежь и современные информационные технологии: сборник трудов XVI Международной научно- практической конференции студентов, аспирантов и молодых ученых, Томск, 3-7 Декабря 2018. - Томск: ТПУ, 2019 - С. 204-205

4. Кашеева Е. В. Выявление зависимости между наличием резистентности к противотуберкулезным лекарственным средствам и предикторами пациента // Сборник избранных статей научной сессии ТУСУР: в 2 частях. Часть 2, Томск, 22-24 Мая 2019. - Томск: В-Спектр, 2019 - С. 97-100

5. Кашеева Е. В. Определение важности признаков пациентов для оценки вероятности резистентности к противотуберкулезным лекарственным средствам // Перспективы развития фундаментальных наук: сборник научных трудов XVI Международной конференции студентов, аспирантов и молодых ученых: в 7 т., Томск, 23-26 Апреля 2019. - Томск: ТПУ, 2019 - Т. 7. IT-технологии и электроника - С. 74-76

6. Кашеева Е. В. Извлечение значимых диагностических критериев для рожистых воспалений из истории болезни // Информационные технологии в науке, управлении, социальной сфере и медицине : сборник научных трудов VI Международной научной конференции , Томск, 14-19 Октября 2019. - Томск: ТПУ, 2019 - С. 31-33

## Список использованных источников и литературы

1. Корнев, М.С. История понятия «большие данные» (Big Data): словари, научная и деловая периодика [Электронный ресурс] / М.С. Корнев. — Электрон. журн. — Москва: Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский государственный гуманитарный университет», 2018. — URL: <https://cyberleninka.ru/article/n/istoriya-ponyatiya-bolshie-dannye-big-data-slovari-nauchnaya-i-delovaya-periodika> (дата обращения 26.05.2020).
2. Big Data [Электронный ресурс] / IT-Enterprise. — Электрон. текстовые дан. — 2018. — URL: <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye> (дата обращения 26.05.2020).
3. Забайкин, А.В. Функция токенизации текста на Python [Электронный ресурс] / Заметки, идеи и скрипты. — URL: <http://zabaykin.ru/> (дата обращения: 10.12.2019).
4. Word embeddings: exploration, explanation, and exploitation (with code in Python) [electronic resource]. — URL: <https://towardsdatascience.com/word-embeddings-exploration-explanation-and-exploitation-with-code-in-python-5dac99d5d795> - Accessed 26.05.2020.
5. TF-IDF [Электронный ресурс] / Википедия. — URL: <https://ru.wikipedia.org/wiki/TF-IDF> (дата обращения: 12.12.2019).
6. Черкасов, В. Л. Роза. Руководство по внутренним болезням: том Инфекционные болезни / под ред. В. И. Покровского. — М., 1996. С. 135–150.
7. Топ-5 языков для машинного обучения [Электронный ресурс] / С. Стельмах. — Электрон. текстовые дан. — 2019. — URL: <https://www.itweek.ru/ai/article/detail.php?ID=207784> (дата обращения 26.05.2020).
8. Криницина, З. В. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение [Электронный ресурс]: учебно-методическое пособие / З. В. Криницина, И. Г. Видяев. — Томск: Изд-во Томского политехнического

университета, 2014. – 73 с. – URL: [https://studopedia.ru/11\\_133767\\_krinitcina-zv-vidyaev](https://studopedia.ru/11_133767_krinitcina-zv-vidyaev) (дата обращения: 25.05.2020).

9. Диаграмма Исикавы [Электронный ресурс]: сайт. – URL: <http://www.up-pro.ru/encyclopedia/diagramma-isikavy.h> (дата обращения: 25.05.2020).

10. История диаграммы Ганта [Электронный ресурс] / Юлия Челянова и Евгений Пикулев. — Электрон. текстовые дан. — Режим доступа: [http://gibtech.ru/blog/discus?entry\\_id=177](http://gibtech.ru/blog/discus?entry_id=177) (дата обращения: 25.05.2020).

11. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 27.12.2018).

12. Федеральный закон от 28 декабря 2013 г. N 426-ФЗ «О специальной оценке условий труда».

13. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» (с изменениями на 21 июня 2016 года).

14. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.

15. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений.

16. СанПиН 2.2.4.3359–16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах.

17. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95.

18. ГОСТ Р 55090-2012 Ресурсосбережение. Обращение с отходами. Рекомендации по утилизации отходов бумаги.

19. ГОСТ Р 55102-2012 Ресурсосбережение. Обращение с отходами. Руководство по безопасному сбору, хранению, транспортированию и разборке отработавшего электротехнического и электронного оборудования, за исключением ртутьсодержащих устройств и приборов.

20. Постановление Правительства РФ от 03.09.2010 № 681 (ред. от 01.10.2013) Об утверждении Правил обращения с отходами производства и потребления в части осветительных устройств, электрических ламп, ненадлежащие сбор, накопление, использование, обезвреживание, транспортирование и размещение которых может повлечь причинение вреда жизни, здоровью граждан, вреда животным, растениям и окружающей среде.

21. НПБ 105-03. Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности.

22. СП 5.13130.2009. Системы противопожарной защиты. Установки пожарной сигнализации и пожаротушения автоматические. Нормы и правила проектирования (с Изменением № 1).

23. СП 3.13130.2009. Системы противопожарной защиты. Система оповещения и управления эвакуацией людей при пожаре. Требования пожарной безопасности.

24. ГОСТ 12.1.030–81 ССБТ Защитное заземление, зануление.

25. ГОСТ Р 12.1.019-2009 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.

26. Безопасность жизнедеятельности. Расчёт искусственного освещения. Методические указания к выполнению индивидуальных заданий для студентов дневного и заочного обучения всех направлений и специальностей ТПУ. – Томск: Изд. ТПУ, 2008. – 20 с.

**Приложение А**  
(справочное)

**Big Data Analytics in Medicine**

Студент

Группа	ФИО	Подпись	Дата
8ПМ8И	Кашеева Евгения Викторовна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Аксёнов Сергей Владимирович	к.т.н.		

Консультант – лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Пичугова Инна Леонидовна	-		

## **Big Data Features**

The concept of “Big Data” has become widespread recently, thanks to the editor of the journal “Nature”, Clifford Lynch, in 2008. Five years later, the following definition of this concept was presented in the Oxford English Dictionary: “sets of information that are too large or too complex to handle, analyse or use with standard methods” [1].

Many areas of human activity have already come across this concept. Medical, educational, financial and other institutions generate large amounts of information. Unfortunately, most of them are not able to work with such data correctly so far.

The McKinsey Global Institute has identified 11 methods and techniques applicable to big data, which in particular include machine learning and visualization of analytic data [2]. Hereafter machine learning is understood as the construction of algorithms that can be taught using mathematical statistics, numerical methods, optimization methods, probability theory, graph theory and various methods of working with digital data. Visualization of analytical data is a representation of the initial information and the results of data analysis in the most convenient for perception and interpretation form.

In addition to the large size, a feature of big data is the data presentation method. In most cases, data is presented in an unstructured or semi-structured way. An unstructured way of presenting data implies the absence of a specific structure; a semi-structured way does not correspond to a strict structure. However, it contains tags and other markers for separating semantic elements [3].

## **Medical Data Formats**

Medical institutions accumulate vast amounts of information. The following patient data is subject to accounting [4]:

- Data collected during the registry or examination of the patient done by the attending physician;
- Tomography and X-ray images;
- Biopsy indications;
- Results of clinical tests;
- Medical appointments, including drugs prescription;
- Biometric data with patient-worn devices and etc.

The data obtained as a result of X-ray and tomography examinations are in DICOM (Digital Imaging and Communications in Medicine) format. Medical documents, in most cases, are text information written in natural language, stored in paper and electronic formats.

To automate their activities, many medical institutions are developing information systems whose functionality is limited to entering data and generating documents and reports based on these data.

We can expand the capabilities of medical information systems using methods and techniques applicable to big data. However, one of the significant problems of using machine learning in medicine is the preparation of the correct medical data for the algorithms, since this requires a large amount of time for specialists of a narrow field.

## Text Data Preprocessing

Working with text data has its own features. The text that a person reads is perceived differently by the machine. The unit of text analysis is not a word, but a lexeme, i.e. a string of encoded bytes representing the text. A string of encoded bytes does not reflect the semantics of the word, does not indicate the relationship and dependence of words. Therefore, without preliminary preparation of the text, the application of machine learning methods is not correct. The stages of preparing the text for analysis include tokenization, stemming (lemmatization), the exclusion of stop words, as well as encoding words into numerical values.

Tokenization refers to the breakdown of the text into smaller parts, tokens, they include words and punctuation marks [5]. Stop words are words that do not carry any semantic meaning.

Stemming is the process of finding the basis of a word, taking into account the morphology of the original word. Stemming performs morphological analysis of a word; it finds a common basis for all grammatical forms of a word, cutting off suffixes and word endings. The most famous stemmer is the Porter Stemmer. This algorithm was written by Martin Porter for common Indo-European languages, including Russian language.

The main idea of Stemmer Porter is that there is a limited number of form-and word-forming suffixes, and word stemming occurs without using any base word basics. Many existing suffixes and manually defined rules are used. The algorithm consists of five steps. At each step, the form-forming or word-forming suffix is cut off and the remaining part is checked for compliance with the rules (for example, for Russian words, the base must contain at least one vowel). If the received word satisfies the rules, the next step is taken. If not, the algorithm selects a different suffix for clipping.

According to the official website of the project, the maximum form-forming suffix is cut off at the first step, the letter “i” is cut off at the second, the word-forming suffix is cut off at the third, and suffixes of excellent forms, “myagkiy znak” and one of the two “n” are cut off at the fourth.

The fact that the Porter Stemmer does not use any dictionaries and bases word basics is a plus for speed and spectrum of application (it copes well with nonexistent words), and at the same time a minus in terms of the accuracy of highlighting the stemm.

Lemmatization is the process of reducing a word form to a lemma – its normal (vocabulary) form. In Russian, the following morphological forms are considered normal forms:

- for nouns this is a nominative case, singular;
- for adjectives this is a nominative case, singular, masculine;
- for verbs, participles this is a verb in an infinitive of imperfect form.

Compared to stemming, the lemmatization algorithm does not delete a part of the word, but replaces the grammatical ending with a suffix or the end of the initial form.

Within the framework of this research project, work is carried out with medical documents “Patient Examination by the Attending Physician”. We tested the work of both methods of obtaining the basis of the word using the example sentences from this document. Table 1 represents a comparison of the operation of the stemmer and the lemmatizer.

Table 1. Results of stemming and lemmatization

Original sentence								
povysheniye	temperatury	s	oznobom	zatem	poyavleniye	giperemii	na	kozhe
Result of stemming								
povyshen	temperatur	s	oznob	zat	poyavlen	giperem	na	kozha
Result of lemmatization								
povysheniye	temperatura	s	oznob	zatem	poyavleniye	giperemiya	na	kozha

We can see that the best option for finding the basis of a word is a lemmatizer. Since when deleting parts of a word, different words can have the same stemming result, which will subsequently affect the analysis results.

Currently, there are a large number of methods for formatting text data to a form suitable for input to machine learning algorithms. The easiest way to encode

words is to number them. However, the numerical data obtained as a result of this method does not carry any semantic load. From the numbers of words it is impossible to determine how close in meaning they are to each other.

There is also a way to reflect the presence or absence of a word in the text, called “One-hot encoding”. The essence of this method is to create a matrix, the columns of which are words, and the rows are documents. In each specific cell of this matrix, either 0 or the number of repetitions of the word in the document is written [6].

Repetition of the same word in several documents indicates that the given word cannot be very important and does not carry much information. To solve this problem, it is necessary to determine the weight of a word using a statistical measure called “TF-IDF” [6, 7].

According to this method, the weight is calculated for each word, the importance of the word is evaluated within a separate document. An assessment of the importance of a  $t_i$  word within a single document is calculated using the following formula:

$$tf(t, d) = \frac{n_t}{\sum_k n_k}, \quad (1)$$

where  $n_t$  is number of repetitions of a  $t$  word in a document and fraction denominator is total number of words in this document.

The inversion of the frequency with which a certain word may be in the collection documents is calculated using the following formula:

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D | t \in d_i\}|}, \quad (2)$$

where  $|D|$  is total number of documents,  $|\{d_i \in D | t \in d_i\}|$  is number of  $D$  collection documents which have a  $t$  word (when  $n_t \neq 0$ ).

The choice of the base of the logarithm in the formula does not matter, since changing the base leads to a change in the weight of each word by a constant factor, which does not affect the ratio of weights. The “TF-IDF” measure is the multiplication of the following two factors:

$$tf-idf(t, d, D) = tf(t, d) \times idf(t, D) \quad (3)$$

Words with a high frequency within a specific document and with a low frequency of use in other documents will gain a lot of weight in “TF-IDF”. This measure is often used in the tasks of information retrieval and text analysis.

For this work, the “TF-IDF” method is most suitable, since there is a need to assess the importance of words in the context of not only the frequent use of the word.

Figure 1 shows the result of encoding using the “TF-IDF” method.



	tf
giperemiya	0.391
golen	0.426
zhar	0.453
levyi	0.426
oznob	0.355
otech	0.296

Figure 1. Result of “TF-IDF” encoding

We see that the word “zhar” is most important for this sentence. Subsequently, the obtained values will be sent to the input of the machine learning algorithm.

## Machine Learning Methods

Supervised Learning, Unsupervised Learning, Reinforcement Learning are the three main types of machine learning. Figure 2 shows a diagram of the types and applications of machine learning algorithms.

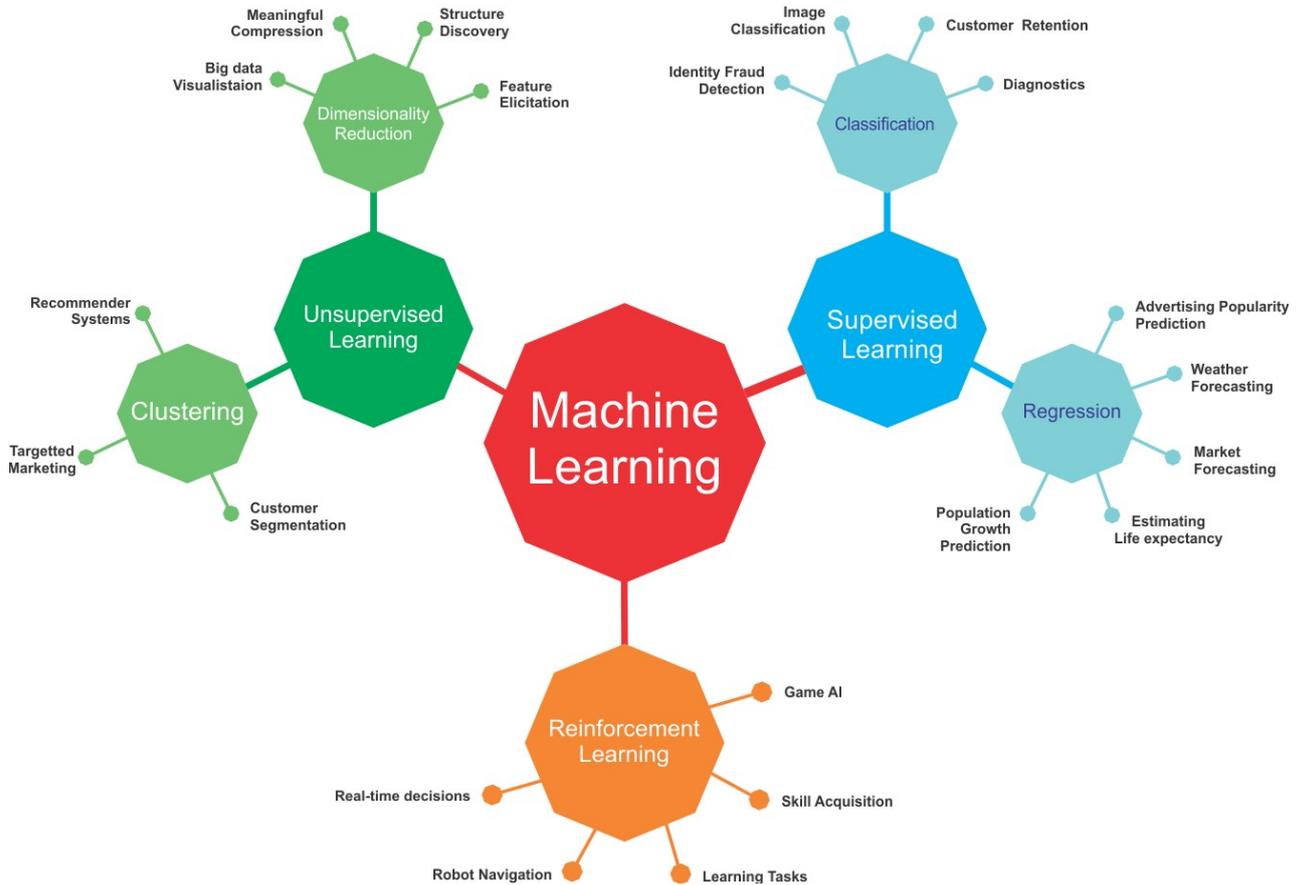


Figure 2. Machine learning methods [8]

The main unsupervised learning algorithm is clustering. Clustering is the task of breaking down multiple objects into groups called clusters. Inside each groups there should be “similar” objects, and objects of different groups should be as different as possible. The list of groups is not clearly defined and is determined during the operation of the algorithm.

The application of cluster analysis in general is reduced to the following steps:

- selection of a sample of clustering objects;
- determination of the set of variables by which objects in the sample will be evaluated;

- calculation of values of measure of similarity between objects;
- application of the cluster analysis method to create groups of similar objects;

- presentation of analysis results.

Supervised learning solves the problems of classification and regression. Regression is used in cases where the results of the algorithm are real numbers or vectors. Classification is applied when the many possible answers are finite. The stages of constructing a classification model are as follows:

- selection of a sample of classification objects;
- class label definition;
- definition of many variables;
- classification method;
- assessment of the quality of the model.

The tasks of medical diagnosis are solved using classification. Having accumulated a sufficient number of precedents in electronic form, it is possible to solve the following problems:

- to classify the type of disease;
- to determine the most appropriate treatment;
- to predict the duration and outcome of the disease;
- to assess the risk of complications;
- to find syndromes, in other words, find the most characteristic set of symptoms for a given disease.

## Classification algorithms

There are many ways to classify objects using machine learning.

The decision tree is one of the most frequently and widely used supervised machine learning algorithms that can perform both regression and classification tasks.

For each attribute in the data set, the decision tree algorithm forms a node where the most important attribute is located in the root node. For evaluation, we start from the root node and go down the tree, following the corresponding node that matches our condition or “decision”. This process continues until a final node is reached that contains the prediction or result of the decision tree.

Let us consider the scenario where a person asks you to lend him a car for one day, and you must decide whether to lend him a car or not. There are several factors that help determine your decision, some of which are listed below:

1. Is this person a close friend or just an acquaintance? If the person is just an acquaintance, then decline the request; if the person is a friend, then go to the next step.
2. Is the person asking for the car for the first time? If so, lend him a car, otherwise go to the next step.
3. Was the car damaged the last time he returned the car? If yes, reject the request; if not, lend him a car.

The decision tree for the above scenario is shown in Figure 3.

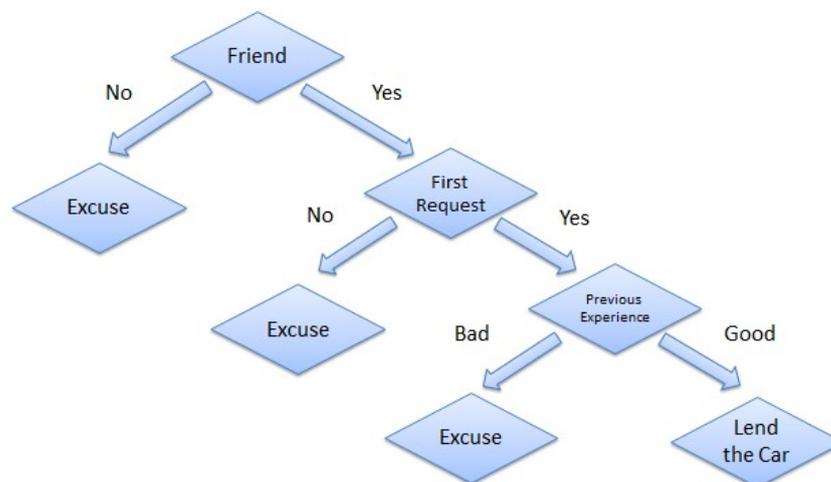


Figure 3. Decision Tree Example [9]

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of training in which you combine different types of algorithms or the same algorithm for several times to form a more powerful forecasting model. In a random forest, an algorithm combines several algorithms of the same type, that is, several decision trees, resulting in a forest of trees, hence the name “Random Forest”. The principle of operation of this algorithm is presented in Figure 4.

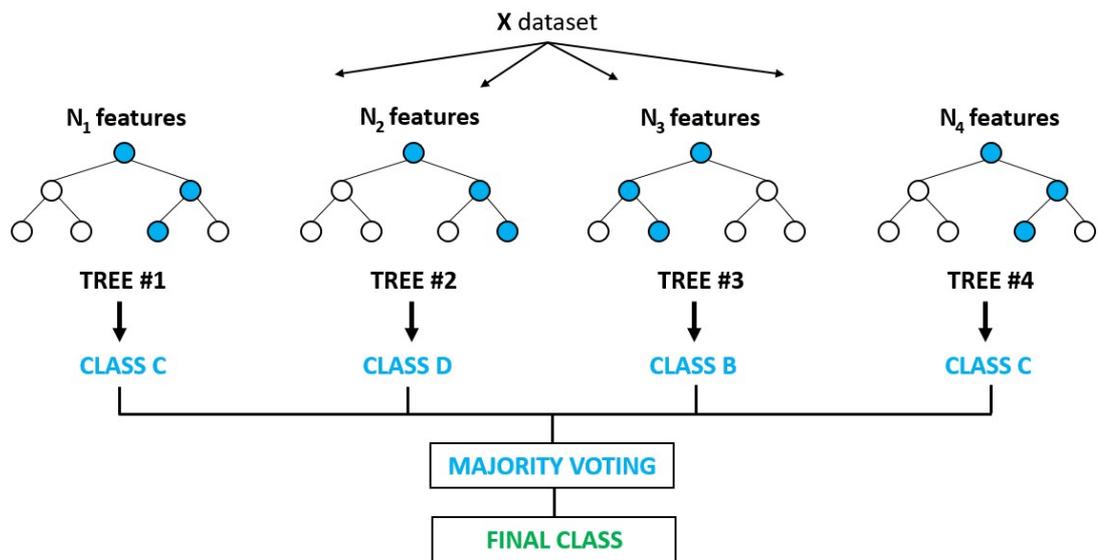


Figure 4. Random Forest Algorithm [10]

The main stages of the random forest algorithm execution are as follows:

1. To select  $N$  random entries from the dataset.
2. To build a decision tree based on these  $N$  records.
3. To select the number of trees in your algorithm and repeat steps 1 and 2.
4. In the case of classification problems, every tree in the forest predicts the category to which belongs to a new record. As a result, a new entry is assigned to the category that receives the most votes.

Like Random Forest, Gradient Boosting is another technique for performing supervised machine learning tasks. Like Random Forests, Gradient Boosting is an ensemble learner. This means that it will create the final model based on a collection of individual models. The predictive power of these individual models is weak and prone to overfitting, but combining many of these weak models into an ensemble will lead to an overall much better result. In Gradient Boosting machines, the most

common type of weak model is decision trees. The principle of operation of this algorithm is presented in Figure 5.

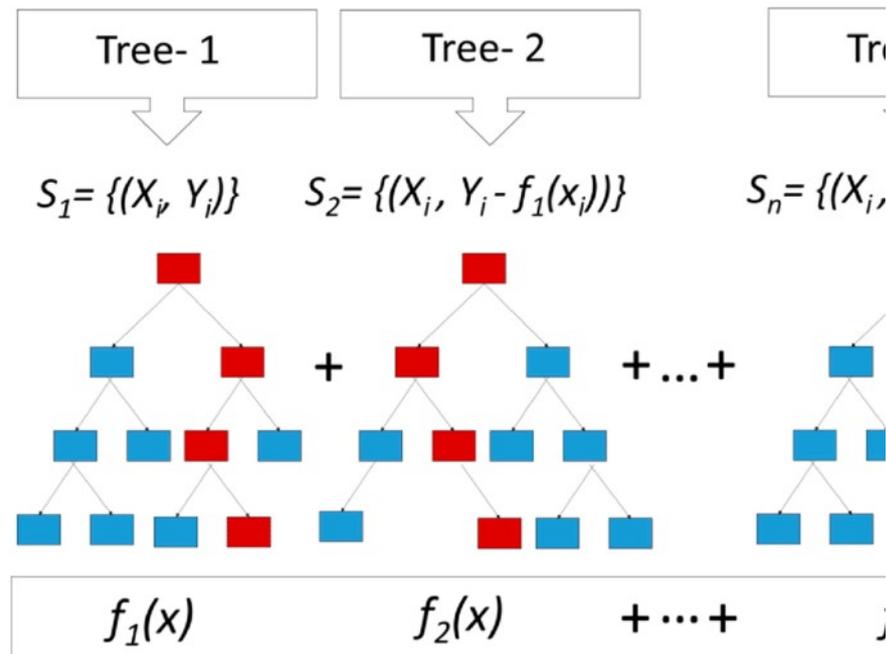


Figure 5. Gradient Boosting Algorithm

Boosting builds models from separate so-called “weak learners” in an iterative way. Separate models are not built on completely random subsets of data and functions, but on a consistent basis, giving more weight to instances with incorrect predictions and high errors. The general idea is that cases that are hard to predict correctly (“difficult” cases) will be concentrated during learning, so that the model learns from the mistakes of the past. When we train each ensemble on a subset of the training set, we also call this Stochastic Gradient Boost, which can help improve the generalizability of our model.

The gradient is used to minimize the loss function. At each stage of training, a weak learner is created, and its predictions are compared with the correct result that we expect. The distance between prediction and truth represents the error rate of our model. These errors can now be used to calculate the gradient. There is nothing special about the gradient, it is basically a partial derivative of our loss function – therefore, it describes the steepness of our error function. The gradient can be used to determine the direction in which you need to change the model parameters in order to (maximize) reduce the error in the next round of training by “descending the gradient”.

## **Conclusion**

Features of the application of big data analysis methods in medicine have been considered in this part of the work.

It was found that the features of big data are large amounts of data, as well as methods for working with this data. Medicine is one of the areas that generates large amounts of information. Since the main part of medical documents is text documents, it was decided to consider the features of working with text data.

Text pre-processing includes tokenization, lemmatization, removal of stop words and formatting of the text into numerical values. Having analysed the existing methods it was revealed that the most suitable method for formatting text into numerical values is the “TF-IDF” method. The essence of this method is that each word is associated with its weight in relation to the text in which it occurs and in relation to all texts in the collection.

After data pre-processing, we can begin to analyze them. There are several machine learning methods, including clustering, classification, regression, and others. For medical diagnosis, classification is used more often. This is due to the fact that this machine learning method allows, on the basis of existing data, to train a model to classify the type of disease, determine the most appropriate treatment, and predict the duration and outcome of the disease.

## References

1. Oxford Learner's Dictionaries [electronic resource]. – URL: <https://www.oxfordlearnersdictionaries.com/definition>. – Accessed 27.05.2020.
2. Big data: The next frontier for innovation, competition, and productivity [electronic resource]. – URL: <https://www.mckinsey.com/business-functions/mckinsey->. – Accessed 27.05.2020.
3. Difference between Structured, Semi-structured and Unstructured data [electronic resource]. – URL: <https://www.geeksforgeeks.org/difference-between-structured-semi-structured-and-unstructured-data/>. – Accessed 27.05.2020.
4. Hightech [electronic resource]. – URL: <https://hightech.fm/2018/09/21/bigdata-med> – Accessed 27.05.2020.
5. Lexical analysis [electronic resource]. – URL: [https://en.wikipedia.org/wiki/Lexical\\_analysis#Tokeni..](https://en.wikipedia.org/wiki/Lexical_analysis#Tokeni..) – Accessed 27.05.2020.
6. Word embeddings: exploration, explanation, and exploitation (with code in Python) [electronic resource]. – URL: <https://towardsdatascience.com/word-embeddings-explor..> – Accessed 27.05.2020.
7. TF-IDF [electronic resource]. – URL: <https://en.wikipedia.org/wiki/Tf-idf> – Accessed 27.05.2020.
8. Machine learning algorithms in simple language. Part 1 [electronic resource]. – URL: <https://medium.com/nuances-of-programming/алгоритмы-м..> – Accessed 27.05.2020.
9. Decision Trees in Python with Scikit-Learn [electronic resource]. – URL: <https://stackabuse.com/decision-trees-in-python-with-scikit-learn/>. – Accessed 4.06.2020.
10. Random Forest Classifier – Machine Learning [electronic resource]. – URL: <https://www.globalsoftwaresupport.com/random-forest-classifier/>. Accessed 5.06.2020.
11. Schematic diagram of a tree-based gradient boosting method [electronic resource] – URL: [https://www.researchgate.net/figure/Schematic-diagram-of-a-tree-based-gradient-boosting-method\\_fig3\\_339077244](https://www.researchgate.net/figure/Schematic-diagram-of-a-tree-based-gradient-boosting-method_fig3_339077244). – Accessed 3.06.2020.

## Приложение Б – Пример документа «Осмотр пациента лечащим врачом»

Номер пациента, пол и возраст  
64, женский, 69

Дата и время осмотра  
21.10.2016, 18:27

### Жалобы

На повышение температуры до 39,2 С, распирающую боль в правом плече и предплечье, чувство жара, тошнота, общая слабость, разбитость, головная боль.

### Анамнез болезни

Заболела остро 21.10.2016г, озноб, повышение температуры до 39,2С, головная боль, разбитость, тошнота, общая слабость. К вечеру заметила эритему правого плеча, распирающую боль, отёк мягких тканей, эритема быстро распространилась на предплечье до лучезапястного сустава. Подобные явления наблюдались неоднократно с 2004г, последний рецидив был в 2012г. Обратилась в СМП, доставлена в инфекционную клинику СибГМУ, госпитализирована.

### Анамнез жизни

В детстве перенесла ветряную оспу, корь, частые ангины, хр. гайморит, хр. бронхит, хр. холицистит, панкреатит, на описторхоз не обследована, рыбу семейства карповых употребляет. СД II типа, глюкозу на уровне 7,5 м/моль/л поддерживает диетой. ГБ, II ст., риск 2, ИБС, стенокардия напряжение ФК II. Хрипение головного мозга. В 1993г прооперирована по поводу аденокарциномы матки. В 2002г операция мастэктомии правой молочной железы в связи со злокачественным новообразованием. Хр. лимфостаз правой в. конечности с 2003г, рецидивирующая рожа с 2004г. Аллергические реакции на плесневой грибок, пыль, медикаменты переносит хорошо. Варикозная болезнь.

### Анамнез ВТЭ

Инвалидность: II группы, в л/н не нуждается.

### Объективный статус

Состояние средней тяжести. Температура 39,2С, вес- 73 кг, рост-160 см, АД - 125/70 мм рт.ст. ЧСС106 в мин. Сознание полное. Кожные покровы и слизистые физиологической окраски, яркая гиперимия лица, разлитая гиперимия зева, зернистость дужек.

### Локальный статус

Эритема на правой руке распространяется от плеча до лучезапястного сустава, горячая на ощупь, границы чёткие, отёк равномерный "тестоватой" консистенции, выраженный. Периферические лимфатические узлы не увеличены, подвижны, безболезненны при пальпации. Подкожно-жировой слой развит удовлетворительно. Мышцы развиты соответственно возрасту, тонус и сила снижены с обеих сторон. Кости и суставы без видимых деформаций, движения в полном объёме, безболезненны. Язык влажный, обложен белым налётом у корня. Щитовидная железа удалена. Грудная клетка цилиндрической формы, пальпация безболезненна. Перкуторный звук лёгочный с обеих сторон, голосовое дрожание равномерное. При аускультации дыхание везикулярное, хрипов нет. Границы сердца расширены влево, тоны приглушены, экстрасистолия, акцент II тона над аортой, тахикардия. Живот обычной формы, равномерно участвует в дыхании, при поверхностной пальпации мягкий, печень у края рёберной дуги, пузырьные симптомы(+). Селезёнка не

увеличена. Кишечник спокоен, эластичен, безболезнен, стул регулярный, оформлен. Симптом "поколачивания" отрицательный с обеих сторон. Диурез в норме. Менингеальных знаков нет.

Диагноз при поступлении

A46 Рожа

Обоснование диагноза

Диагноз поставлен на основании острого начала, повышения температуры до 39,2С, синдром интоксикации выражен, предшествовал характерным местным явлениям. Рожа правого плеча и предплечья рецидивирующая, хр. лимфостаз правой в., конечности. Заболеванию предшествовало переохлаждение.

Диагноз

A46 Рожа правого плеча и предплечья, распространённая, эритематозная, рецидивирующая, средней степени тяжести.

**Приложение В – Пример документа «Температурный лист»**

Дата	22.10		23.10		24.10		25.10		26.10		27.10		28.10		29.10		30.10	
	У	В	У	В	У	В	У	В	У	В	У	В	У	В	У	В	У	В
Температура	36,2	37,3	36,6	36,3	36	36,5	36	36,5	36	36,3	36,5	36,4	36,5	36,5	36,4	36,5	36,5	36,4
Давление систолическое	14		144		128		110		120		135				130			
Давление диастолическое	75		78		69		70		80		80				90			
Пульс	68		63		64		69		70		72				76			