

Школа Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.03.01 «Информатика и вычислительная техника»  
 Отделение школы Отделение информационных технологий

### БАКАЛАВРСКАЯ РАБОТА

Тема работы
<b>Методы анализа данных для прогнозной аналитики в нефтегазовой компании</b>

УДК 004.415.2:004.421:004.65:622.276

Студент

Группа	ФИО	Подпись	Дата
3-8В71	Бондаренко Антон Викторович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Ботыгин Игорь Александрович	к.т.н., доцент		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Кашук Ирина Вадимовна	к.т.н., доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
старший преподаватель	Авдеева Ирина Ивановна	–		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Погребной А.В.	к.т.н.		

Томск – 2022 г.

**Запланированные результаты обучения выпускника образовательной программы бакалавриата, указанными в ФГОС ВПО по направлению 09.03.01 «Информатика и вычислительная техника»**

Код результата	Результат обучения (выпускник должен быть готов)	Требования ФГОС, критерии АИОР
Р1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности.	Требования ФГОС (ОК-7, ОПК-5, ПК-3), критерий 5 АИОР (п. 1.1)
Р2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.	Требования ФГОС (ОК-7, ОПК-2,5, ПК-1, 3), критерий 5 АИОР (п. 1.1, 1.2)
Р3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей.	Требования ФГОС (ОК-6, ОПК-1, ПК-2, 4), критерий 5 АИОР (п. 1.2)
Р4	Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т.п.) в соответствии с техническим заданием и с использованием средства автоматизации проектирования.	Требования ФГОС (ОК-7, ОПК-2, 4, ПК-1, 2), критерий 5 АИОР (п. 1.3)
Р5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем.	Требования ФГОС (ОК-5, ОПК-5, ПК-1, 2, 3), критерий 5 АИОР (п. 1.4)
Р6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.	Требования ФГОС (ОК-8, 9, ОПК-1, 2, 4, ПК-3, 4, 5), критерий 5 АИОР (п. 1.5)
	<i>Универсальные компетенции</i>	
Р7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.	Требования ФГОС (ОК-3, ОПК-3, 5),

		критерий 5 АИОР (п. 2.1)
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты, комплексной инженерной деятельности	Требования ФГОС (ОК-5, 7, ПК-3, 4), критерий 5 АИОР (п. 2.2)
P9	Эффективно работать индивидуально и в качестве члены группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.	Требования ФГОС (ОК- 2, 6, 7), критерий 5 АИОР (п. 2.3, 2.4)
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 2, 3, 4, 5), критерий 5 АИОР (п. 2.5)
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.	Требования ФГОС (ОК-5, 7), критерий 5 АИОР (п. 2.6)

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

Школа: Инженерная школа информационных технологий и робототехники  
 Направление подготовки: 09.03.01 «Информатика и вычислительная техника»  
 Отделение: Отделение информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП  
 \_\_\_\_\_ Погребной А.В.  
 (Подпись)    (Дата)    (Ф.И.О.)

**ЗАДАНИЕ  
на выполнение выпускной квалификационной работы**

В форме:

Бакалаврской работы

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

Студенту:

Группа	ФИО
3-8В71	Бондаренко Антон Викторович

Тема работы:

Методы анализа данных для прогнозной аналитики в нефтегазовой компании	
Утверждена приказом директора (дата, номер)	№ 40-48/с от 09.02.2022

Срок сдачи студентом выполненной работы:	11.06.2022
--	------------

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b></p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Объектом исследования является анализ данных в нефтегазовой компании на основе методов машинного обучения для прогнозирования отказов насосно-компрессорных труб.</p> <p>Среда разработки программного обеспечения Jupyter Notebook. Язык программирования Python.</p>
---	---

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>  <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>	<ol style="list-style-type: none"> <li>1. Обзор предметной области;</li> <li>2. Разработка информационной системы прогнозирования;</li> <li>3. Обработка и анализ данных с использованием машинного обучения;</li> <li>4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение;</li> <li>5. Социальная ответственность.</li> </ol>
<p><b>Перечень графического материала</b>  <i>(с точным указанием обязательных чертежей)</i></p>	<ol style="list-style-type: none"> <li>1. Информация по нефтегазовому комплексу;</li> <li>2. Информация по машинному обучению;</li> <li>3. Матрица корреляций.</li> </ol>

<p><b>Консультанты по разделам выпускной квалификационной работы</b>  <i>(с указанием разделов)</i></p>	
<p><b>Раздел</b></p>	<p><b>Консультант</b></p>
<p>Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</p>	<p>Кацук Ирина Вадимовна</p>
<p>Социальная ответственность</p>	<p>Авдеева Ирина Ивановна</p>

<p><b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b></p>	<p>24.01.2022</p>
--	-------------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Ботыгин Игорь Александрович	к.т.н., доцент		24.01.2022

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
3-8В71	Бондаренко Антон Викторович		24.01.2022

## РЕФЕРАТ

Выпускная квалификационная работа содержит 134 с., 58 рис., 24 табл., 49 источников, 1 прил.

Ключевые слова: анализ данных, машинное обучение, искусственные нейронные сети, Data Mining, насосно-компрессорные трубы.

Объектом исследования является отказ насосно-компрессорных труб в процессе эксплуатации.

Цель работы – исследование методов анализа данных в нефтегазовой компании на основе методов машинного обучения для прогнозирования отказов насосно-компрессорных труб.

В процессе исследования проводились обработка и анализ данных с использованием методов машинного обучения.

В результате исследованы методы анализа данных для прогнозной аналитики в нефтегазовой компании на основе методов машинного обучения для прогнозирования отказов насосно-компрессорных труб.

Степень внедрения: проводится тестирование.

Область применения: нефтегазовая отрасль.

Экономическая эффективность/значимость работы: снижение затрат на ремонт и обслуживание нефтепромыслового оборудования.

В будущем планируется применение результатов работы в производственной деятельности.

## **Сокращения, обозначения, термины, нормативные ссылки**

НГК – нефтегазовый комплекс.

НГО – нефтегазовая отрасль.

НКТ – насосно-компрессорная труба.

ИС – информационная система.

ПА – предиктивная аналитика.

МО – машинное обучение.

ИИ – искусственный интеллект.

ИНС – искусственная нейронная сеть.

ПО – программное обеспечение.

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

Школа Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.03.01 Информатика и вычислительная техника  
 Отделение школы Отделение информационных технологий

Форма представления работы:

Бакалаврская работа
---------------------

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

Срок сдачи студентом выполненной работы:	11.06.2021
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный Балл раздела (модуля)
22.02.2022	Объект и методы исследования	10
07.03.2022	Мониторинг состояния и анализ данных в нефтегазовой сфере деятельности	10
28.03.2022	Разработка интеллектуальной информационной системы прогнозирования отказов насосно-компрессорных труб нефтяных и газовых скважин	20
02.05.2022	Анализ данных отказов насосно-компрессорных труб нефтяных и газовых скважин с использованием методов машинного обучения	40
25.05.2022	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
08.06.2022	Социальная ответственность	10

**СОСТАВИЛ:****Руководитель ВКР**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИТ	Ботыгин Игорь Александрович	к.т.н., доцент		24.01.2022

**СОГЛАСОВАНО:****Руководитель ООП**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИТ	Погребной Александр Владимирович	к.т.н.		24.01.2022

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

Студенту:

<b>Группа</b>	<b>ФИО</b>
3-8В71	Бондаренко Антон Викторович

<b>Школа</b>	Инженерная школа информационных технологий и робототехники	<b>Отделение</b>	<b>Отделение информационных технологий</b>
<b>Уровень образования</b>	Бакалавриат	<b>Направление</b>	09.03.01 Информатика и вычислительная техника

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

<i>1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Стоимость материальных ресурсов и специального оборудования определены в соответствии с рыночными ценами г. Томска Тарифные ставки исполнителей определены штатным расписанием НИ ТПУ
<i>2. Нормы и нормативы расходования ресурсов</i>	Норма амортизационных отчислений на специальное оборудование
<i>3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Отчисления во внебюджетные фонды 30 %

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

<i>1. Анализ конкурентных технических решений (НИ)</i>	Расчет конкурентоспособности SWOT-анализ
<i>2. Формирование плана и графика разработки и внедрения (НИ)</i>	Структура работ. Определение трудоемкости. Разработка графика проведения исследования
<i>3. Составление бюджета инженерного проекта (НИ)</i>	Расчет бюджетной стоимости НИ
<i>4. Оценка ресурсной, финансовой, бюджетной эффективности (НИ)</i>	Интегральный финансовый показатель.  Интегральный показатель ресурсоэффективности.  Интегральный показатель эффективности.

**Перечень графического материала**

1. Оценка конкурентоспособности ИП
2. Матрица SWOT

- |   |
|---|
| 3. Диаграмма Ганта<br>4. Бюджет НИ<br>5. Основные показатели эффективности НИ |
|---|

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

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Кащук Ирина Вадимовна	к.т.н доцент		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
3-8В71	Бондаренко Антон Викторович		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа		Ф.И.О.	
3-8В71		Бондаренко Антон Викторович	
Школа	Инженерная школа информационных технологий и робототехники	Отделение (НОЦ)	Отделение информационных технологий
Уровень образования	бакалавриат	Направление/специальность	09.03.01 Информатика и вычислительная техника

Тема ВКР:

<i>Анализ данных для прогнозирования показателей в нефтегазовой компании</i>	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
<b>Введение</b> —	<p><i>Объект исследования:</i> анализ данных в нефтегазовой компании на основе методов машинного обучения для прогнозирования отказов насосно-компрессорных труб</p> <p><i>Область применения:</i> нефтегазовая промышленность</p> <p><i>Рабочая зона:</i> офис</p> <p><i>Размеры помещения:</i> 5х6 м.</p> <p><i>Количество и наименование оборудования рабочей зоны:</i> ПЭВМ – 1 шт.</p> <p><i>Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне:</i> работа с компьютерным оборудованием</p>
<b>Перечень вопросов, подлежащих исследованию, проектированию и разработке:</b>	
<b>1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения:</b> —	<p>Нормативные документы, регламентирующие организацию трудового процесса на рабочем месте:</p> <ol style="list-style-type: none"> <li>«Трудовой кодекс Российской Федерации» от 30.12.2001 N 197-ФЗ (ред. от 25.02.2022) (с изм. и доп., вступ. в силу с 01.03.2022);</li> <li>ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования»;</li> <li>ГОСТ 21889-76 «Система «Человек-машина». Кресло человека-оператора. Общие эргономические требования».</li> <li>Федеральный закон от 28 декабря 2013 г. N 426-ФЗ «О специальной оценке условий труда»</li> </ol>
<b>2. Производственная безопасность при разработке проектного решения:</b> —	<p><b>Опасные факторы:</b></p> <ol style="list-style-type: none"> <li>Анализ электробезопасности (удар электрическим током, короткое замыкание, статическое электричество).</li> </ol> <p><b>Вредные факторы:</b></p> <ol style="list-style-type: none"> <li>Анализ освещенности рабочей зоны;</li> <li>Анализ уровня шума и вибрации;</li> <li>Анализ показателей микроклимата.</li> </ol> <p><b>Требуемые средства коллективной и индивидуальной защиты от выявленных факторов:</b> беруши, наушники, защитные ограждения, устройства защитного отключения, дополнительные источники освещения и средства нормализации микроклимата.</p>
<b>3. Экологическая безопасность при разработке проектного решения</b>	<p><b>Воздействие на селитебную зону:</b> не происходит</p> <p><b>Воздействие на литосферу:</b> бумага, средства печати, перегоревшие люминесцентные лампы, вышедшие из строя компоненты оргтехники.</p> <p><b>Воздействие на гидросферу:</b> продукты жизнедеятельности персонала</p> <p><b>Воздействие на атмосферу:</b> не происходит</p>

<p><b>4. Безопасность в чрезвычайных ситуациях при разработке проектного решения</b></p>	<p><b>Возможные ЧС:</b>  природные катастрофы (экстремально низкие и высокие температуры воздуха, наводнения, цунами, ураган, лесные пожары и т.д.);  геологические воздействия (землетрясения, оползни, обвалы, провалы территории и т.д.);  техногенные (террористическая деятельность, аварии на электро-, тепло-коммуникациях, водоканале и т.д.)  <b>Наиболее типичная ЧС:</b> пожар в здании.</p>
--	---

Дата выдачи задания для раздела по линейному графику

**Задание выдал консультант:**

Должность	ФИО (полностью)	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Авдеева Ирина Ивановна			

**Задание принял к исполнению студент:**

Группа	ФИО (полностью)	Подпись	Дата
3-8В71	Бондаренко Антон Викторович		

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	15
1. МОНИТОРИНГ СОСТОЯНИЯ И АНАЛИЗ ДАННЫХ В НЕФТЕГАЗОВОЙ СФЕРЕ ДЕЯТЕЛЬНОСТИ .....	17
1.1 Анализ актуальности и проблематики нефтегазовой сферы деятельности.....	17
1.2 Анализ возможностей и преимуществ использования современных методов интеллектуального анализа данных .....	23
1.3 Анализ особенностей прогнозирования данных .....	27
1.4 Методы, модели, функционирование нейросетей.....	32
2. РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ОТКАЗОВ НАСОСНО-КОМПРЕССОРНЫХ ТРУБ НЕФТЯНЫХ И ГАЗОВЫХ СКВАЖИН .....	42
2.1 Анализ существующих программных продуктов и систем проведения интеллектуального анализа данных.....	42
2.2 Анализ ключевых преимуществ программных библиотек поддержки процессов анализа данных .....	50
2.3 Построение UML-диаграмм для отображения статической структуры системы, поведенческих и физических аспектов системы .....	55
2.4 Обобщённая функциональная структура системы прогнозирования .....	61
3. ПРОВЕДЕНИЕ ОБРАБОТКИ И АНАЛИЗА ДАННЫХ ОТКАЗОВ НАСОСНО-КОМПРЕССОРНЫХ ТРУБ НЕФТЯНЫХ И ГАЗОВЫХ СКВАЖИН С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ .....	67
3.1 Формализация концепции проводимых исследований .....	67
3.2 Описание набора данных для анализа.....	68
3.3 Проведение обработки и анализа данных с использованием методов машинного обучения 73	
4. Раздел «Финансовый менеджмент» .....	92
5. Раздел «Социальная ответственность» .....	111
ЗАКЛЮЧЕНИЕ .....	124
СПИСОК ЛИТЕРАТУРЫ.....	125
ПРИЛОЖЕНИЕ А. ЛИСТИНГ КОДА .....	129

## ВВЕДЕНИЕ

Нефтяная промышленность в настоящее время является основным сектором топливно-энергетического комплекса нашей страны. От его успешного функционирования зависят эффективное удовлетворение внутреннего и внешнего спроса на нефть и продукты ее переработки, обеспечение валютных и налоговых поступлений в государственный бюджет. Не менее важна эта отрасль и для энергетической безопасности страны и ее политических интересов в мире [1]. В связи с этим, важной задачей является проведения комплексного и регулярного анализа целевых данных для повышения эффективности функционирования всех составляющих системы нефтяной промышленности, что связано с рядом проблем.

Ключевая сложность аналитики данных в рассматриваемой предметной области состоит в их высоком разнообразии, гетерогенности, не структурированности, зашумленности, избыточности

Развитие методов обработки данных в нефтегазовой отрасли, во многом, сдерживается как технологическими особенностями, так и ограничениями существующей инфраструктуры сбора и подготовки данных, отсутствием эффективных платформ аналитики больших данных, позволяющих собрать и контекстуализировать разнородные данные, реализовать современные аналитические алгоритмы. Для решения подобных задач целесообразным является внедрение методов и моделей предиктивной аналитики (ПА)

Конечная цель ПА заключается в выработке стратегии для учета возможных рисков, поиск потенциала для роста, определение новых возможностей и прогнозирование будущих событий. Для этого используются статистические методы экономического анализа, математические модели, методы интеллектуального анализа данных, теории игр, устойчивые шаблоны поведения (паттерны) и многое другое [2].

Использование ПА в контексте рассматриваемой нефтегазовой отрасли необходимо в первую очередь для:

- анализа и прогнозирования влияния факторов на параметры добычи нефтегазовых продуктов;
- прогнозирования отказов оборудования нефтегазовой промышленности, в том числе для перехода от обслуживания по регламенту к обслуживанию по состоянию;
- прогнозирования объемов добычи продукции, потребления энергии и других материальных ресурсов.

Особенную актуальность ПА приобретает в сфере технического обслуживания и ремонта нефтедобывающего и перерабатывающего сырья оборудования, для чего применяются средства машинного обучения (МО) и искусственного интеллекта (ИИ). В отличие от базовых механизмов контроля, установленных производителями оборудования, они имеют более широкий потенциал, позволяющий анализировать и визуализировать дополнительные факторы, влияющие на состояние оборудования. Все это дает возможность прогнозировать сценарии возникновения отказов оборудования, предупредить фатальные ошибки и, как следствие, сократить расходы.

Целью настоящей работы является исследование методов анализа данных в нефтегазовой компании на основе методов машинного обучения для прогнозирования отказов насосно-компрессорных труб в процессе эксплуатации.

Задачами работы являются:

- Мониторинг состояния и анализ данных в нефтегазовой сфере деятельности.
- Разработка интеллектуальной информационной системы прогнозирования отказов насосно-компрессорных труб.
- Проведение обработки и анализа данных отказов насосно-компрессорных труб нефтегазовой отрасли с использованием методов машинного обучения.

# 1. МОНИТОРИНГ СОСТОЯНИЯ И АНАЛИЗ ДАННЫХ В НЕФТЕГАЗОВОЙ СФЕРЕ ДЕЯТЕЛЬНОСТИ

## 1.1 Анализ актуальности и проблематики нефтегазовой сферы деятельности

Современный нефтегазовый комплекс (НГК) является одной из наиболее важных составляющих общей топливно-энергетической базы всего государства. Стратегия развития НГК во многом определяет основополагающие особенности нефтегазовой отрасли, влияя на различные прогнозные показатели и принятие целевых решений по повышению уровня энергетической независимости государства на мировом рынке энергетики и его конкурентоспособности.

Нефтегазовая промышленность государства в настоящее время считается одной из ключевых сфер формирования отечественной экономики, определяющей эффективность и рентабельность функционирования других прикладных отраслей, косвенно влияя и на уровень общего благосостояния нашего населения [1].

Наибольшую актуальность приобретают задачи, связанные с проектированием и реализацией эффективных методов повышения целевых показателей нефте- и газодобычи, что связано с ограниченностью объемов данных ресурсов и сложностью проведения работ по их возобновлению.

Мировой НГК представляет собой сложный комплекс, включающий в себя ряд взаимодействующих субъектов, характеризующихся периодической нестабильностью и высокой зависимостью от различных внешних факторов, в том числе политических решений и природных катаклизмов [3].

Специфика НГК в нашем государстве заключается в том, что газовая, нефтяная и нефтеперерабатывающая сферы промышленности являются наиболее активно развивающимися и доминирующими как в региональных, так в локальных масштабах. В связи с ростом направлений и сфер производственной деятельности в государстве постоянно повышается уровень значимости энергетических

ресурсов, таких как нефть и газ, в том числе для использования в сфере промышленности органического синтеза.

Без подобных сырьевых компонентов промышленные производственные линии, такие как металлургия и машиностроение, не могут эффективно осуществлять свою функциональную деятельность, т.к. являются энергозатратными.

Мировой НГК регулярно совершенствуется, активно внедряя различные инновационные методы, решения и прикладное оборудование, требуемое для обеспечения замкнутого цикла всех производственных мероприятий, направленных в первую очередь на разработку новых месторождений сырья, добычу и переработку нефтегазовых продуктов с последующей доставкой конечному потребителю [4].

В процессе разведывания и поиска новых месторождений сырья, разработки и добычи полезных ископаемых, этапов последовательной сепарации и очистки, приведения в товарный вид с последующей перевозкой на производственные предприятия требуется использование специализированного технического оборудования, обеспечивающего выполнение нормативов и стандартов во всех перечисленных процессах.

Следует отметить, что ключевыми этапами современного освоения месторождений нефти (рис.1.1) являются такие [5]:

1. Открытие новых нефтяных месторождений.
2. Оценка имеющихся в наличии запасов нефтяных месторождений.
3. Технологическая и организационная подготовки к освоению месторождения.
4. Непосредственная добыча нефти и газа из месторождений.
5. Ликвидация месторождения после исчерпания его запасов.

Проблематика НГК в рамках нашей страны обусловлена рядом факторов, ключевой проблемой является исчерпаемость сырья и ограниченность числа месторождений. С каждым годом имеющиеся в НГК запасы нефтяных и газовых залежей стабильно сокращаются, что влечет за собой повышение глубины уровня

их добычи и способствует значительному усложнению условий их производства. Все это приводит к снижению объемов поставки и переработки нефтегазовых ресурсов, что повышает стоимость на многие категории товаров и услуг, напрямую зависящих от цен на топливо, например транспортные перевозки, логистика, аграрный сектор [6].



Рисунок 1.1 – Ключевые этапы освоения нефтяных месторождений

Среди наиболее значимых проблем НГК можно выделить такие [7]:

- активное ежегодное сокращение объемов общей добычи нефтегазовых ресурсов;
- низкая эффективность проводимой диверсификации инфраструктуры снабжения данными ресурсами в различных регионах государства;
- высокий объем потребления добываемого природного газа населением и производственными промышленными организациями;
- высокая зависимость рынка от решений и ценовой политики компаний-монополистов в НГК;

- не эффективное распределение загрузки функционирующих нефтеперерабатывающих заводов;
- значительные объемы не выполненных долговых обязательств в нефтегазовом секторе со стороны потребителей.

При этом особенно следует отметить, что нефтегазовая отрасль (НГО) в настоящее время характеризуется сложностью влияния на человеческий фактор, в частности высокой степенью монополизации, не эффективной политикой менеджмента и принятия управленческих решений, низким уровнем прозрачности процессов труда, слабой конкуренцией [8]. Все это порождает ряд рисков для всей отрасли в целом (рис.1.2).

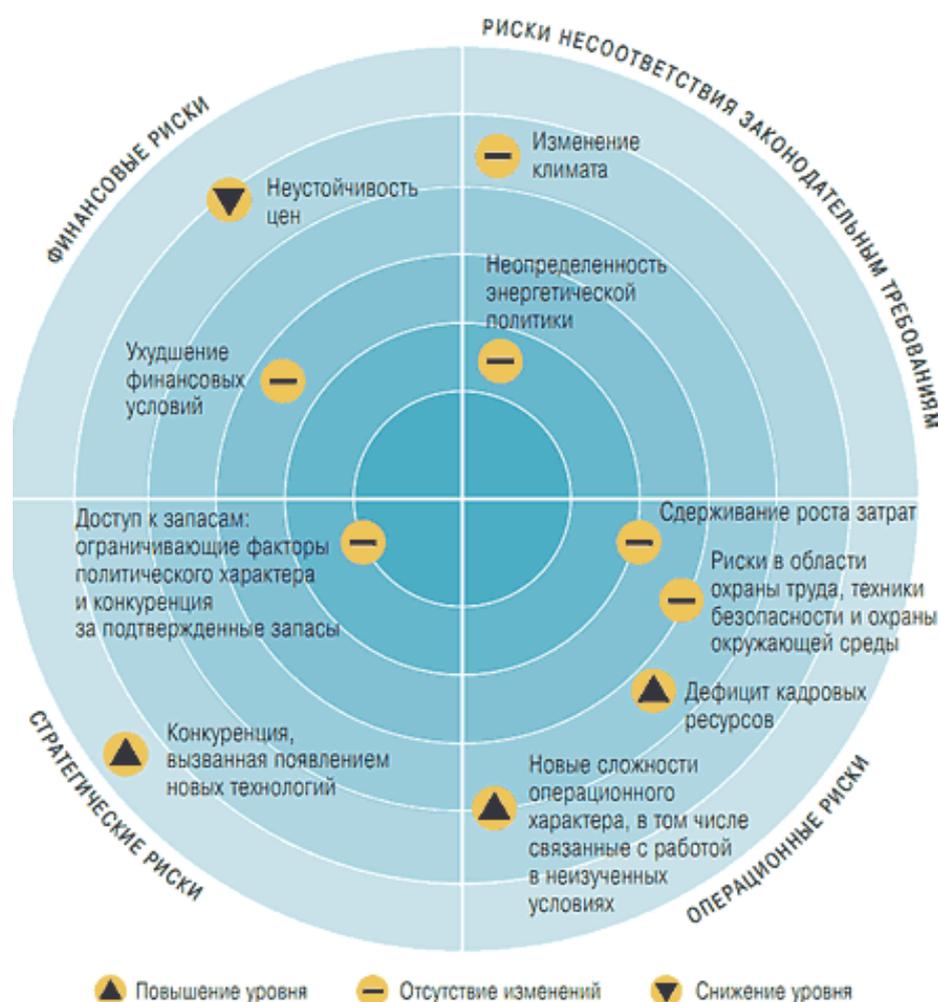


Рисунок 1.2 – Ключевые риски НГО

В отличие от нефти, газ при его добыче и подготовке не требует значительных временных и финансовых затрат по переработке перед

непосредственным использованием, однако сроки его поставки потребителю должны быть максимально короткими. В связи с этим возникают определенные особенности в организации логистических процессов и транспортировки ресурсов.

Переработка добываемой нефти и производство нефтепродуктов выполняется на специализированных нефтеперерабатывающих заводах, многие из которых на сегодняшний день является устаревшими по технологическим уровням и не могут эффективно конкурировать с европейскими аналогами [9].

Проведя анализ ряда литературных источников [3-6] следует отметить, что среди проблем современных добывающих предприятий НГО нашей страны наиболее часто выделяются геологические, технологические и экономические виды (рис.1.3).

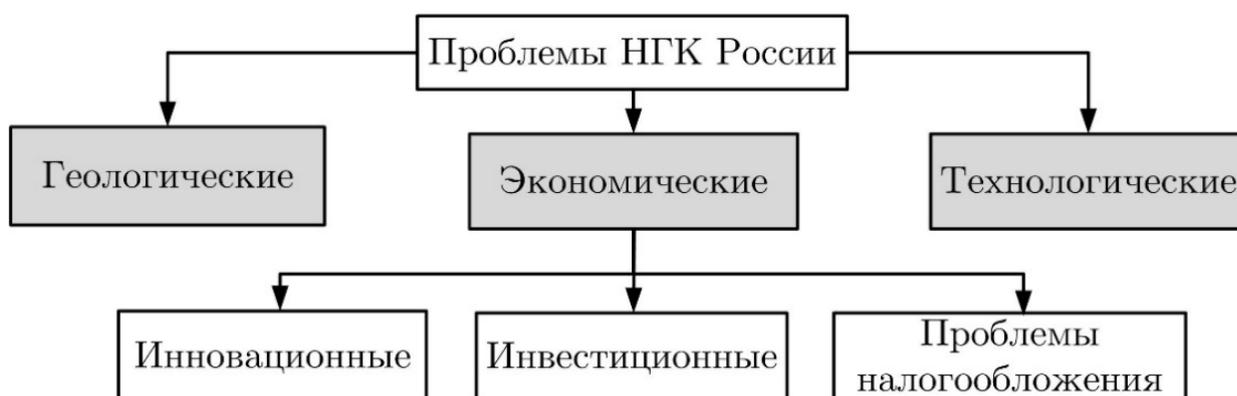


Рисунок 1.3 - Классификация проблем НГО России

Среди проблем первого типа НГО (*геологических*) уже на этапе добычи углеводородов ключевыми проблемами являются естественное истощение недр и доступных запасов ресурсов, что приводит к появлению большого количества труднодоступных залежей, которые проблематичны из-за тяжелых условий добычи (сложные природно-климатические или горно-геологические условия) и по причине особенностей их химических свойств. В частности, для нефти, содержащей различные примеси при ее добыче, требуется использование специализированного оборудования и применения апробированных нетрадиционных методов добычи сырья, т.к. нефть находится глубоко в земле [10].

К наиболее значимым типам современных проблемам *технологического* характера в НГК относят значительный уровень износа добывающего технологического оборудования, устаревшие и не эффективные технологии добычи сырья.

Следует отметить, что перечисленные недостатки характерны не только на этапе добычи углеводородов, но и при проведении непосредственных геологоразведочных работ. Специфика состоит в том, что отечественные технологии и общая материально-техническая база менее развиты, чем в других ведущих государствах, поэтому возникает необходимость в приобретении дорогостоящего специализированного иностранного оборудования, что сказывается на себестоимости процессов.

Ключевые *экономические* проблем на конкретных предприятиях НГК, во многом, обусловлены существующим нормами в налогообложении, постоянными модификациями в действующем законодательстве [7]. В частности, основные проблемы инновационного характера в первую очередь связаны с недостаточным уровнем финансирования научно-технических исследований и опытно-конструкторских работ различного профиля. По этой причине повышается уровень затрат на проведение строительных работ по устройству скважин, возрастает себестоимость добываемого сырья [9].

Дополнительное негативное воздействие оказывают проблемы оттока иностранных инвестиций, в связи с тем, что НГО для них становится менее прибыльным вложением из-за санкционных запретов и ограничений. Уменьшение объемов вложения отечественного капитала по причине существующего экономического кризиса, в том числе из-за коррупции и бюрократизацией, также усложняет процесс функционирования НГК.

Анализ рассмотренных аспектов и проблематики в НГК позволяет выделить следующие, наиболее вероятные сценарии их решения:

- расширение типов и методов проведения геологоразведочных работ для открытия новых месторождений сырья, в том числе под водой;

- создание резервной базы полезных ископаемых непосредственно в районах их добычи;
- разработка и использование новейших технологий по добыче нефтегазового сырья, направленных на повышение уровня нефтеотдачи пластов посредством комплексной интенсификации разработки залежей природных ресурсов;
- интеграция экономических реформ, в том числе имплементация гибкой системы налогообложения, разделенного налога в зависимости от уровня прибыльности компании по прогрессивной шкале с учетом льготных периодов и государственной поддержки по определенным критериям;
- сотрудничество и перенимание опыта у зарубежных партнеров;
- использование прогнозной аналитики для превентивной оценки и анализа данных по ресурсам НГК.

Также следует отметить, что проблему повышения объемов добычи полезных ископаемых целесообразно решать посредством общей модернизации технической составляющей НГК.

Таким образом, современная НГО сегодня является одной из наиболее значимых на отечественном рынке сферой, несмотря на активную популяризацию альтернативных источников энергии и конкуренцию между компаниями в этом экономическом сегменте. И одним из самых перспективных направлений в НГО является автоматизация промышленного и аналитического процессов по добыче сырья, обеспечивающая снижение временных и материальных затрат по ведению эффективной производственной деятельности организациями данного профиля.

## 1.2 Анализ возможностей и преимуществ использования современных методов интеллектуального анализа данных

Ключевым направлением ИИ, используемым для решения задач распознавания образов, является машинное обучение (МО). Теория МО возникла в один период с созданием персональных компьютеров (ПК) и весь период до

сегодняшнего дня активно совершенствуется и дополняется. Актуальность МО заключается в том, что его использование позволяет автоматизировать процессы обработки и анализа данных посредством задействования вычислительных возможностей современных компьютеров и использования хранилищ данных больших объемов для их агрегации в единую аналитическую экосистему.

Дополнительной причиной активного применения МО на практике является постоянное увеличение числа прикладных областей, для которых целесообразными задачами являются повышение эффективности использования ресурсов путем построения прогностических моделей разного рода. МО нацелено на реализацию процессов построения математических и вероятностных моделей, оценки их качества работы для проведения комплексного исследования информационных потоков, а также разнородных данных.

Непосредственно само обучение созданных моделей проявляется в виде поиска и конфигурации различных числовых и строковых параметров, весовых коэффициентов, которые можно модифицировать с целью максимизации требуемого уровня качества объекта исследования. Таким образом, формируемая структура обеспечивает пути создания элементов ИИ на основе анализа подаваемых наборов входных данных, фактически, на статистической информации. В результате выполнения всех обозначенных процессов итоговые сформированные модели могут быть использованы для решения задач прогнозирования, кластеризации, регрессии или классификации для выявления новых знаний или скрытых данных [12].

Одной из ключевых задач современного МО является поддержка процессов автоматической идентификации выявленных взаимозависимостей между скрытыми (неочевидными) массивами переменных (признаков) объекта исследования и явными, что может быть эффективно использовано и для решения задач прогнозной аналитики компании. Это может быть полезным в случаях, когда необходимо провести оценку вероятных значений скрытых компонентов сформированного объекта по явным данным.

На базовом уровне МО можно классифицировать на два ключевых типа (рис. 1.4).

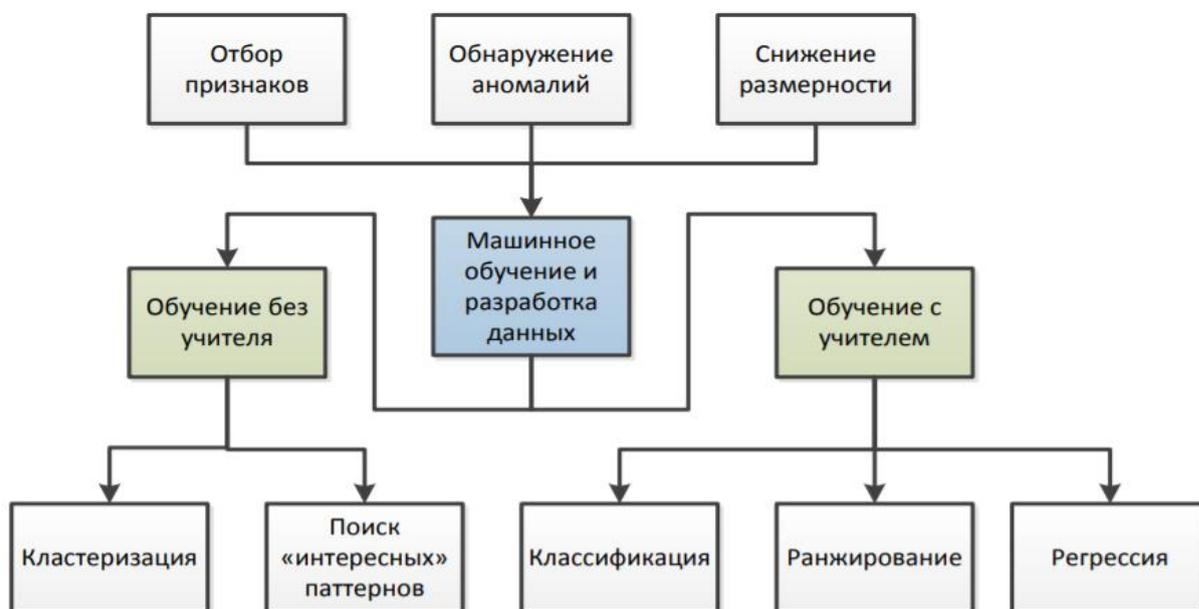


Рисунок 1.4 – Обобщенная схема классификации методов МО

Тип 1. МО с учителем нацелено на построение моделей, позволяющих выводить выходное значение признака по поданной ей на вход последовательности входных данных. После создания объекта модели её необходимо инициализировать и применять для присвоения итоговых меток неизвестным ранее данным. Это является задачей классификации (в случае названий классов или целочисленных значений) и регрессии (вещественные значения). Таким образом, решения задачи классификации – итоговые признаки представляют собой различные числовые категории, а в случае регрессии – являются непрерывными значениями [13].

Тип 2. МО без учителя заключается в несколько другой специфике, т.к. в данном случае целью использования МО является анализ входного набора данных, который заранее не имеет разметки, т.к. выделенных выходных признаков для выявления зависимостей или оценки близости данных между собой. Такие модели чаще всего используются для решения задач снижения размерности данных или кластеризации. Последняя задача наиболее популярна в тех случаях, когда требуется выделить явные группы данных по отдельным признакам. Понижение

размерности является актуальным при сжатии или уменьшении объемов однородных или высоко коррелирующих данных.

Концептуальная схема процесса МО приведена на рис. 1.5.

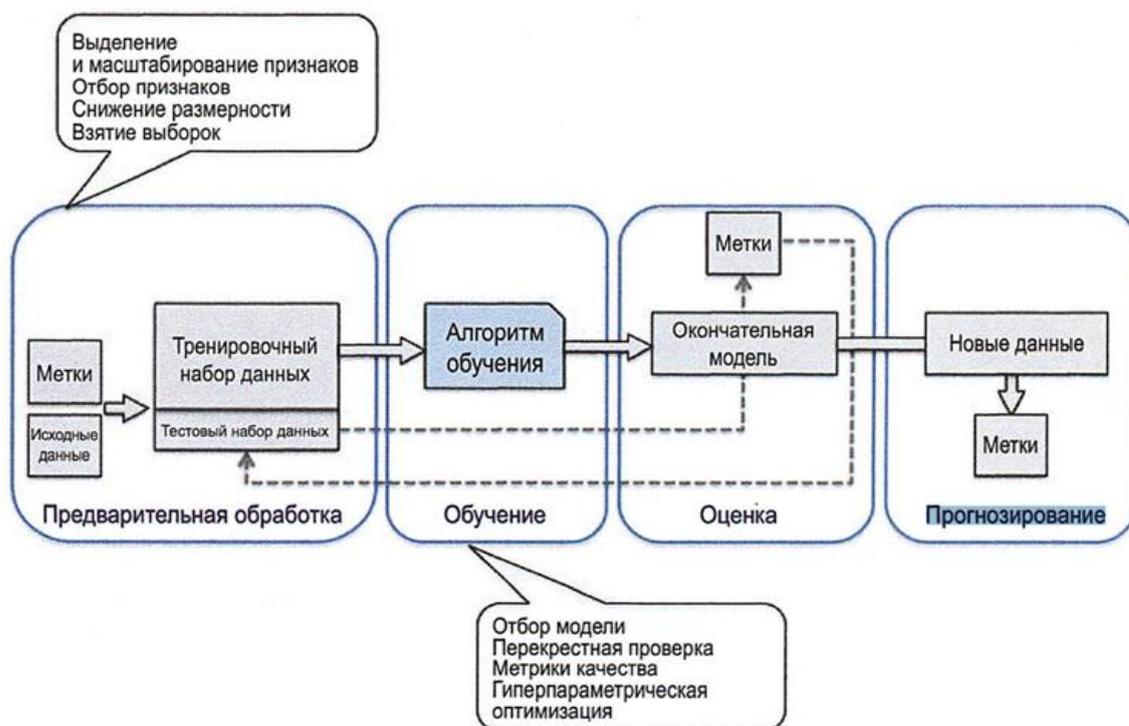


Рисунок 1.5 – Концептуальная схема общего процесса выполнения МО

Преимущества использования МО в прикладных программных системах и приложениях заключаются в следующем [14]:

- Высокий уровень точности работы на больших массивах данных. Благодаря имплементации логики программы на основе анализа входных данных, релевантных конкретной задаче, становится выше обобщающая способность модели в процессе выявления новых данных и численных метрик оценки качества ее работы.

- Системная автоматизация. Благодаря используемому математическому аппарату в МО становится возможным динамическое выявление корреляции, как в данных, так и в самих признаках, что приводит к созданию новых, более гибких, шаблонов. Преимуществом данного подхода является интеграцию таких

автоматизированных систем в рабочие процессы прикладной сферы использования МО [15].

- Быстродействие оптимизированных моделей для анализа. Оптимизированные модели МО способны оперативно обработать набор входных данных для выдачи результатов в онлайн режиме посредством распределения вычислительных процессов по пересчёту весовых значений, что особенно актуально для задач анализа потоковых данных.

- Гибкость конфигурации и сериализации созданных моделей.

- Высокий уровень масштабируемости. На всех этапах расширения рассматриваемых задач МО созданные модели могут быть гибко распределены по микросервисной архитектуре для эффективного использования вычислительных ресурсов в параллельном режиме.

### 1.3 Анализ особенностей прогнозирования данных

С развитием и постоянным ростом алгоритмической и вычислительной сложности аппарата прогнозирования, с все большей производительностью и доступностью компьютеров, имеющих возможности объединения в сетевые кластеры, оснащенных прикладным распределенным программным обеспечением (ПО), прогнозирование становится эффективно реализуемой задачей [21].

Основной целью проведения прогноза является снижение уровня неопределенности, в рамках которого лицо, принимающее решение (ЛПР) должно выбирать необходимую альтернативу. Данная цель формирует 2 следующих правила для проведения процедур прогнозирования:

- процессы должны быть выполнены технически корректным образом, порождая сбалансированные прогнозы с точностью, соответствующей заданным задачам;

- процессы и результаты прогнозирования должны быть наглядными для интерпретации ЛПР с целью использования для обоснования выбора решений прикладных задач.

Прогнозы часто подразделяются на долгосрочные и краткосрочные. Первые требуются для формирования основополагающих курсов или стратегий управления потоками на длительный период. Краткосрочные прогнозы применяются для менее долгосрочных стратегий [22].

Основные аспекты процесса решения задачи прогнозирования приведены на рис.1.6.

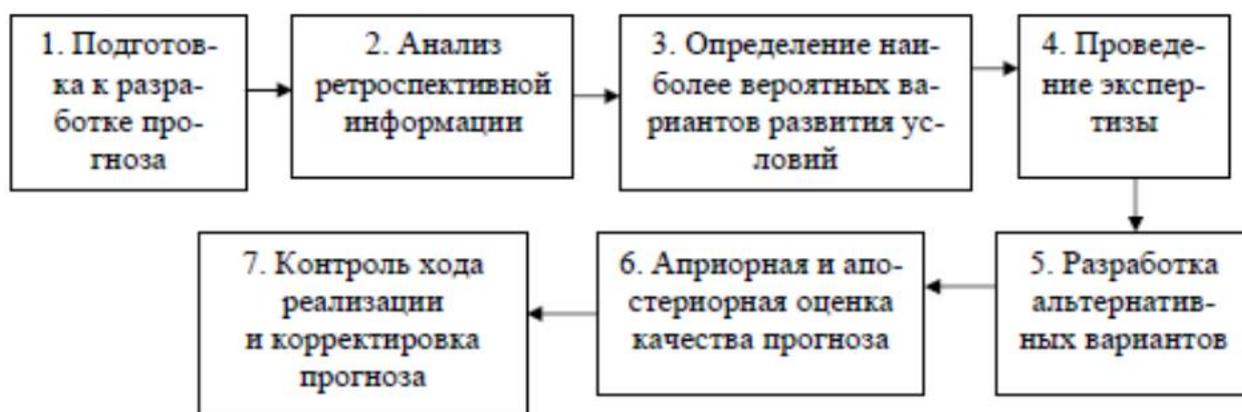


Рисунок 1.6 – Основные аспекты процесса решения задачи прогнозирования

Прогнозы часто подразделяются относительно их отношению к степени детализации учета отдельных составляющих или всех обобщенных показателей работы системы.

Наиболее часто процедуры прогнозирования подразделяются на количественные и качественные.

Формализованные процедуры построения прогнозов основываются на переносе полученного ранее опыта и данных в неопределенное будущее. Т.е., такие процедуры базируются на предположении, что условия, сформировавшие ранее данные, трудно отличимые от будущих условий, с учетом исключений по переменным, распознанных моделью более точно.

Осмысление специфики оперирования данными прогнозирования, которые сформированы естественными событиями, позволяет выделить следующие этапы в процессе прогнозирования:

1. Сбор данных, который предполагает получение верифицированных и валидных данных. Данный этап является наименее точной частью всего процесса

проведения прогнозирования, т.к. более сложен для проведения проверки, в связи с тем, что последующие этапы могут достаточно просто производиться с использованием поданных данных.

2. Уплотнение данных, необходимо для выполнения процессов прогнозирования в случаях, когда собрано избыточное или недостаточное количество данных. Это обусловлено проблемой, заключающейся в том, что некоторые данные часто не имеют непосредственного отношения к рассматриваемой задаче, уменьшая точность прогнозов, а другие данные соответствуют рассматриваемой проблеме, в ограниченном контексте заданного исторического периода.

3. Построение модели прогноза и ее численная оценка. Данный этап заключается в подборе такой модели прогноза, которая наилучшим образом описывает особенности собранных данных с точки зрения снижения ошибки прогноза. При этом на популярность и удобство использования модели влияет уровень ее простоты, от чего зависит доверие к полученному прогнозу у ЛПР.

4. Экстраполяция выбранной модели прогнозирования, что предусматривает получение требуемого прогноза фактическим образом, т.к. необходимые данные являются собранными, редуцированными с определенной моделью прогноза. Для проверки точности формируемых результатов используется прогнозирование на краткосрочные периоды.

5. Оценка значений полученного прогноза. На этом этапе проводится последовательное сравнение вычисленных величин с реально наблюдаемыми значениями. Для этого часть актуальной информации исключается из всего множества анализируемых данных. Таким образом, после подбора модели реализуется прогноз на выделенные периоды, а полученные результаты подвергаются сравнению с уже известными значениями, которые эмпирически доказаны. На данном этапе некоторые процедуры предусматривают суммирование рассчитанных абсолютных значений ошибок, представляя данную сумму или усредненное значение (среднюю ошибку прогноза). Более гибкие процедуры оценки используют сумму квадратов ошибок для дальнейшего сравнения с

аналогичными значениями, полученными для других методов прогнозирования [23].

Общая классификация методов прогнозирования временных рядов приведена на рис.1.7.



Рисунок 1.7 - Общая классификация существующих методов прогнозирования

Методы прогнозирования классифицируют на 2 отдельные группы: интуитивные и формализованные. Первый тип методов опирается на использование суждений и оценок экспертов в рассматриваемой предметной области. Данные подходы популярны в маркетинге, экономике, политике, в виду сложности систем, прогнозирование поведения которых трудоемко с математической точки зрения или, наоборот, является слишком простым.

Формализованные методы определяют скрытую математическую зависимость, используемую для вычисления будущих значений рассматриваемых процессов с целью построения прогноза. К наиболее популярным относятся [21]:

а) *Экспертные методы*. Например, метод интерактивного прогнозирования Дельфи. Его сущность заключается в структурированном сборе мнений различных экспертов для их дальнейшего обобщения в единую формализованную оценку. В

случае, когда исследователю требуется выполнить прогнозирование каких-либо параметров или характеристик финансовых рынков, необходимым является выделение выбранной экспертной группы людей и проведение их параллельного или последовательного анкетирования, или опроса. На базе полученных результатов выполняется обобщение о текущей ситуации.

б) *Методы логического моделирования* – базируются на процедурах поиска и выявления скрытых или не очевидных закономерностей долгосрочной перспективы производства.

К таким методам относятся [23]:

- сценарный метод, используемый для формализации последовательностей результатов происшествия события, путем разработки предметной базы знаний;
- методы аналогий;
- методы прогнозов по образу.

в) *Экономико-математические методы*, использующие в своей основе подходы к созданию моделей изучаемого объекта. Модель данного типа представляет собой некоторую схему, отображающую возможный (наиболее вероятный) путь развития производства по ряду условий. При прогнозировании временных рядов на базе данной категории методов применяются статистические, динамические, микро- и макро-, линейные, нелинейные, и оптимизационные модели, базирующиеся на составлении систем уравнений, в которые вводятся различные ограничения. Ключевую роль играет функционал или критерий оптимальности, представляющий собой специальное уравнение. С его помощью возможно нахождение оптимального прогнозного значения по одному или нескольким показателям [22].

г) *Статистические методы* прогнозирования для временных рядов базируются на формировании различных сводных индексов (диффузный, смешанный), а также на расчете значений математического ожидания, дисперсии, вариации, экстраполяции или интерполяции.

д) *Технический анализ* – представляет собой метод прогнозирования изменений исследуемых параметров в дальнейшем отрезке времени на базе последовательного анализа изменений параметров в прошлом [23].

е) *Фундаментальный анализ* – представляет собой метод прогнозирования, базирующийся на комплексной процедуре анализа производственных и финансовых показателей работы компании. Данный метод часто используется для оценки общей стоимости акций компании, что позволяет отразить рентабельность ее функционирования [22].

ж) *Искусственные нейросети*. С помощью данного метода можно построить сложные модели, особенно для больших наборов данных. Метод чувствителен к масштабированию данных и выбору параметров, большим моделям требуется много времени для обучения.

Одно из их главных преимуществ ИНС заключается в том, что они могут использоваться для обработки разнородной информации, которая характеризуется большими объемами, для формирования сложных прогнозных моделей. В случае, если компания обладает значительными аппаратными ресурсами, входные данные очищены и структурированы, а модель настроена правильно, то данный метод способен существенно превзойти по результатам прогнозирования другие методы и алгоритмы.

В связи с этим сложные и массивные модели ИНС требуют длительного времени обучения, аппаратных ресурсов и предварительной обработки входных данных. ИНС на практике лучше всего работают с однородными данными, когда все признаки измерены в едином масштабе [24].

#### 1.4 Методы, модели, функционирование нейросетей

С точки зрения прогнозирования в рамках рассматриваемой задачи наиболее перспективными являются модели ИНС. Рассмотрим их особенности использования подробнее.

Концептуально, типовая нейронная сеть (НС) может быть отражена в виде набора связанных клеток в человеческом мозгу посредством специальных связей – синапсов. Подобная модель взята за основу при моделировании искусственной НС (ИНС) на вычислительных устройствах для ее адаптации и использования при проведении сводного многопараметрического анализа данных больших объемов. ИНС на практике в вычислительной информационной системе создается на базе агрегации функциональных компонентов, обеспечивающих расчетные операции по оценке весовых коэффициентов при передаче сигналов. Все данные, которыми оперирует созданная модель ИНС, преобразуются и передаются через некоторый набор слоёв ИНС, включающих в свой состав набор специальных искусственных нейронов [25].

Искусственный нейрон (ИН), в свою очередь, необходим в качестве основы модели формирования биологического нейрона с соответствующим поведением и логикой работы. На вход каждому отдельному ИН поступает совокупность сигналов, формируемых на основе работы предыдущих нейронов в сети. Получаемое значение мультиплицируется по каждому входу на заданное весовое число, соответствующее уровню степени силы синапсов. После этого проводится операция суммирования для формирования итогового значения степени активации ИН [26]. На рис.1.8 приведена модель типового ИН.

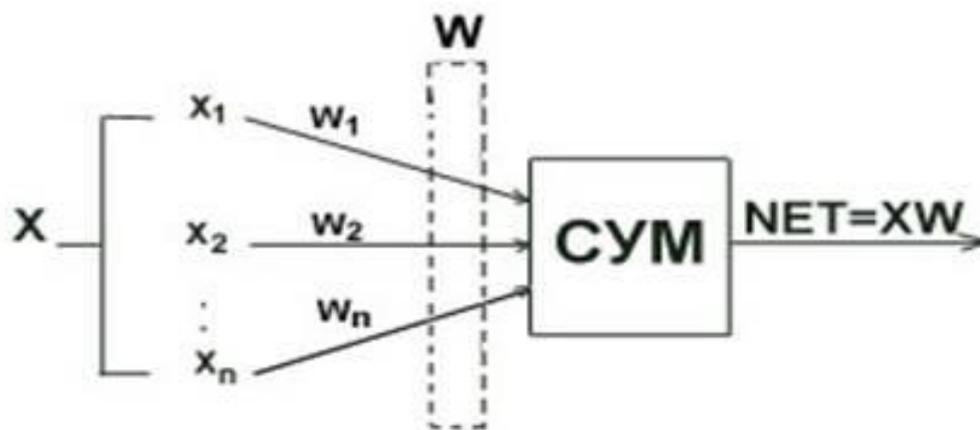


Рисунок 1.8 – Модель искусственного нейрона

Общая схема обучения ИНС на базе алгоритма обратного распространения ошибки приведена на рис.1.9.

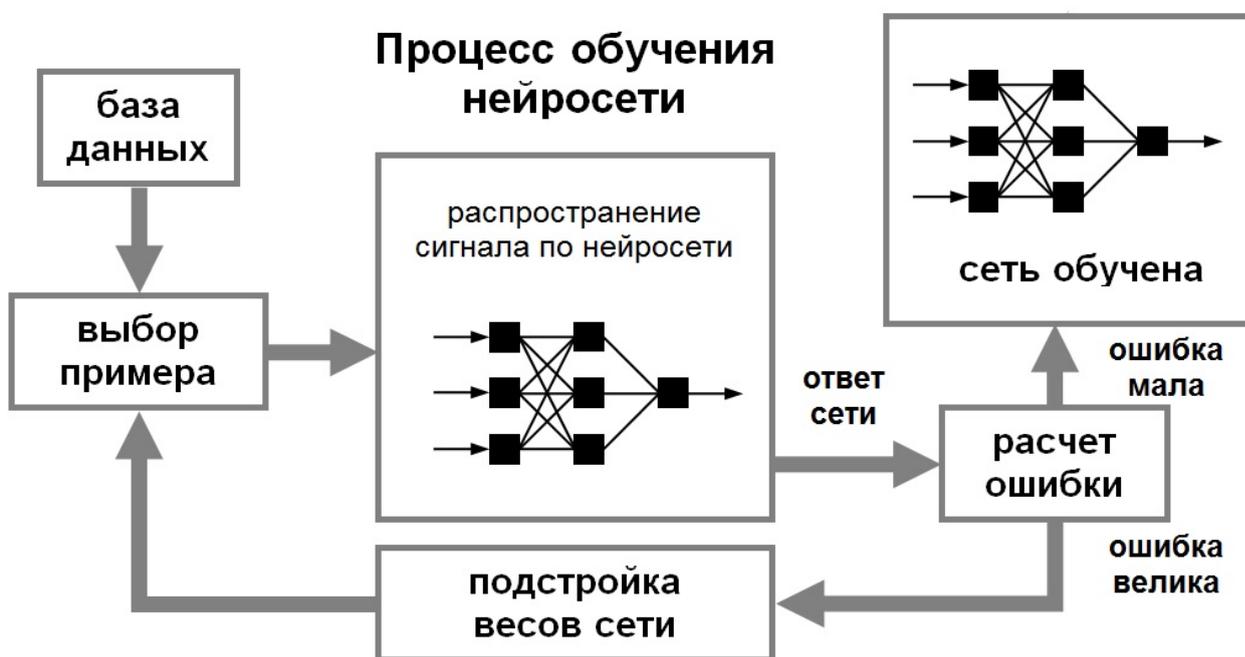


Рисунок 1.9 – Общая схема обучения ИНС

Обучение большинства современных ИНС выражается в форме последовательного вычислительного процесса, в котором используемые связи и гиперпараметры модели ИНС задаются посредством моделирования логики создания узлов модели. Характеристики обучения модели ИНС задаются на основе методов подстройки задаваемых параметров и изменения их значений [27].

Большинство современных моделей ИНС могут быть обучены на основе обработки входных данных по тестовым примерам с размеченными выходными признаками, однако, существуют модели, предназначенные для обработки данных без разметки.

В настоящее время в научно-практическом направлении использования ИНС существует ряд разных архитектур моделей ИНС, которые могут быть адаптированы для различных сценариев и выявления функциональных зависимостей. Классификация классических и современных моделей ИНС в обобщенном виде приведена на рис.1.10.

Чаще всего по параметру направленности связей модели ИНС классифицируются на обратные (СОС) и без использования обратных связей (БОС), т.е. однонаправленные архитектуры [28].

Преимуществами последних типов (БОС) является простота и скорость их программной имплементации с последующим последовательным получением итогового ответа после осуществления операций обучения. Недостатком подобной архитектуры является снижение вычислительного объема операций и размеров модели ИНС, это снижает обобщающую способность модели и может способствовать тому, что каждый отдельный ИН будет многократно оказывать влияние на модель по каждой итерации, понижая итоговую точность прогноза.

К СОС архитектурам относят модели Хопфилда для поиска ассоциаций и Кохонена для кластеризации данных при обучении без учителя. Преимуществами моделей СОС являются большая детализация процесса обучения посредством наличия значительного числа ИН в ИНС, а также высокая точность на хорошо структурированных наборах данных. Недостатками подобных моделей являются необходимость в проведении процессов настройке большого числа взаимосвязанных гиперпараметров, а также требуется соблюдение ряда специальных условий для проведения адекватного обучения модели, при этом требуются достаточно высокие вычислительные мощности.



Рисунок 1.10 – Классификация типовых моделей ИНС

В последние годы все большую актуальность для прогнозирования на практике приобретают рекуррентные и свёрточные ИНС.

Рекуррентными называются ИНС (РИНС). Идея РИНС заключается в порядке процесса обработки данных, в частности, в классических РИНС существует правило, что каждый вход и выход модели являются независимыми друг от друга, в связи с чем сфера применения данных моделей ограничена характером данных и типом задачи. В составе РИНС искусственные нейроны

связаны между собой одной или несколькими обратными связями. Пример подобной архитектуры приведен на рис. 1.11.

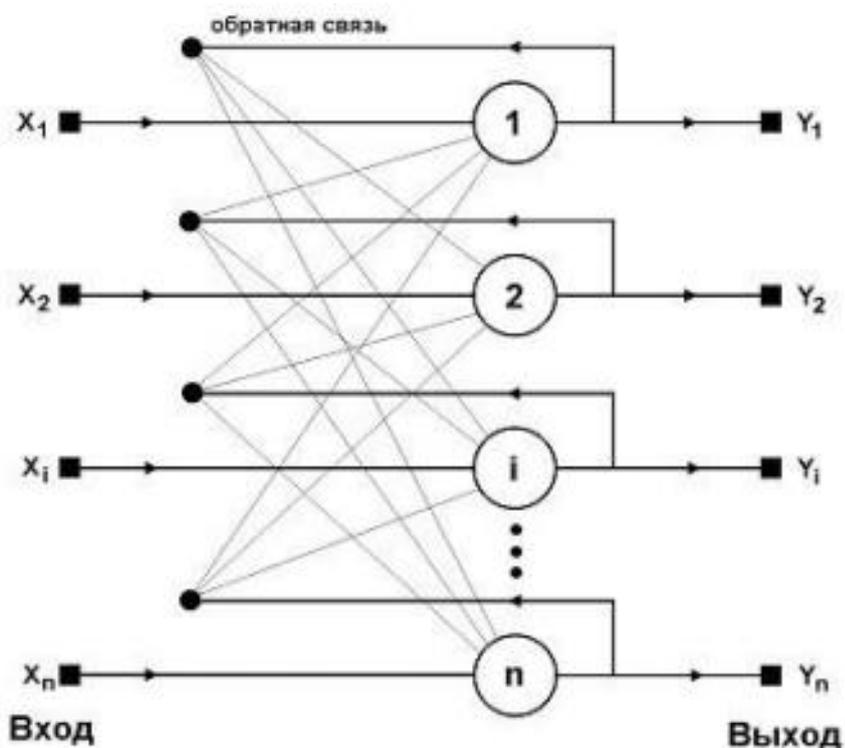


Рисунок 1.11 – Структура типовой РИНС в общем виде

Подобные связи необходимы для обеспечения возможностей использования дополнительных воздействий на элементы модели, поэтому РИНС часто применяются в задачах с необходимостью задействования ассоциативной памяти. Кроме того, такие связи между нейронами моделей позволяют РИНС формировать набор определенных состояний, что удобно для хранения дополнительных метаданных в процессах обработки данных [29].

Активно используемым видом РИНС в практике современного ИАД являются модели с длинной цепью элементов краткосрочной памяти (Long short-term memory, LSTM), обладающие дополнительными возможностями по хранению долгосрочных зависимостей в данных (рис. 1.12).

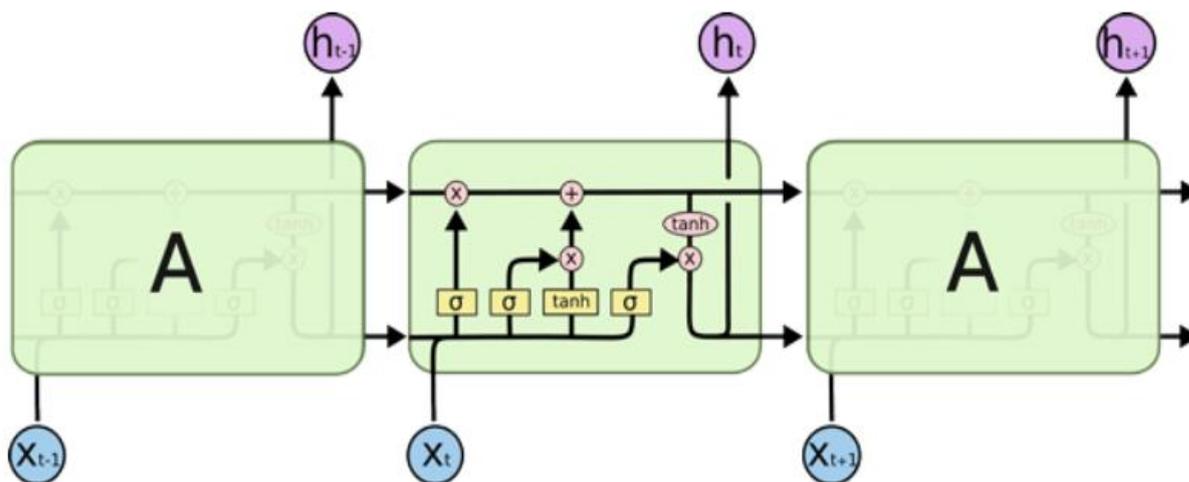


Рисунок 1.12 – Схема построения модели LSTM типа

Такие модели ИНС в своем составе элементов содержат систему ворот, которые в динамическом режиме могут переключать своё текущее состояния под конкретный сигнал, что позволяет реализовать концепцию хранения памяти в долгосрочной форме [30].

Свёрточные модели ИНС (СНС) применяют элементы, которые позволяют проводить модификацию весов и их смещения. Каждый элемент СНС модели получает на вход набор данных, над которыми проводится операция скалярного произведения, после чего, в соответствии с сформированными правилами, осуществляются нелинейные преобразования, в частности, операции свертки.

Архитектура данных ИНС моделей накладывает ряд ограничений на входные данные, что удобно благодаря возможности кодирования заданных свойств под структуру созданной модели для повышения скорости обработки наборов данных. Посредством подобной специфики первоначальная процедура объявления может быть имплементирована более эффективно, что способствует уменьшению числа значимых параметров модели.

СНС модели позволяют агрегировать в себе ряд функциональных принципов, являющихся необходимыми для обеспечения достаточного уровня обобщающей способности, а также борьбы с пространственными искажениями разного рода. Наиболее важными из таких принципов являются:

- локальные рецепторные поля;

- использование пространственных выборок для организации необходимой иерархии;
- синаптические коэффициенты.

Рассматриваемая модель ИНС может включать в свой состав разные типы слоев (векторов ИН), в том числе свёрточные, субдискретизирующие и персептроны. Пример структуры СНС модели приведен на рис.1.13.

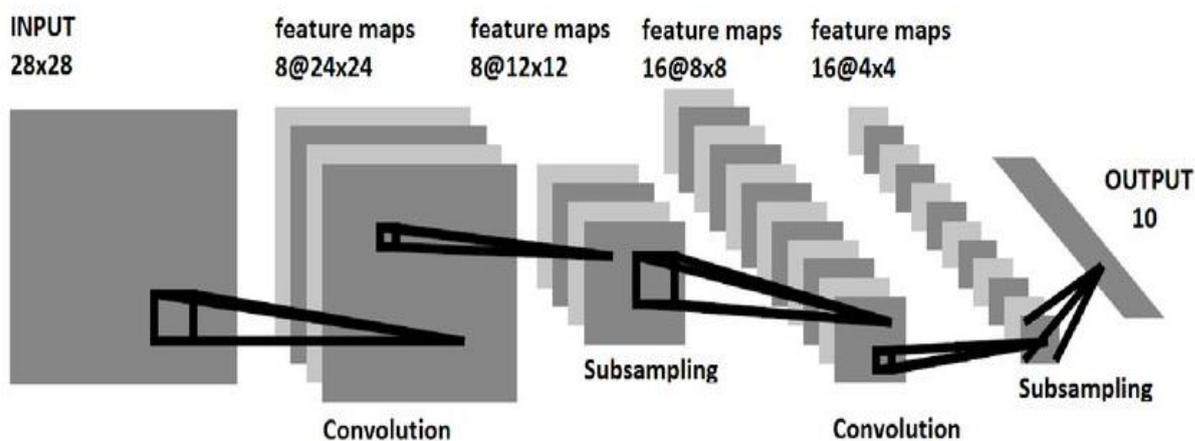


Рисунок 1.13 – Топология структуры СНС модели

Свёрточные и субдискретизирующие слои обеспечивают формирование входного вектора путем чередования между собой для выделения наиболее значимых признаков модели.

Ключевой вычислительной операцией в моделях СНС, важной с точки зрения прогнозирования, является свёртка. Фактически, данный процесс проходит в следующих этапах: ядро ИНС перемещается посредством перехода в двумерных наборах данных, проводя операции умножения на те фрагменты входных данных, в окрестности которых оно расположено в каждый момент времени, с последующей агрегацией в виде суммирования.

Ядро ИНС производит такую процедуру итеративно по каждому фрагменту изображения. Признаки, сформированные на выходе СНС модели, таким образом, представляют собой взвешенные суммы входных признаков, размещенных в том

участке пространства, где находятся выходные значения пикселей входного вектора [31].

Принадлежность признака входному набору может быть выявлена путем оценки вхождения его в текущую зону ядра по входным данным. Фактически, размерность ядра модели СНС позволяет задать итоговое значение признаков, необходимых для получения нового вектора выходного признака. Основным преимуществом данного типа модели является поддержка концепции общих весов, что особенно актуально с точки зрения вычислительных затрат на больших объемах данных в сфере прогнозирования. Благодаря этому фактору, СНС модели могут снижать число гиперпараметров, которые требуется подобрать и установить для проведения процесса обучения.

На практике используются специфические модели СНС, базирующиеся на неоконгнитронных представлениях, посредством чего связанность весовых значений является минимальной. Благодаря подобной архитектуре модели СНС могут быть оперативно использованы как в режиме применения последовательных вычислительных узлов, так и в распределенной системе, организовав процесс свёртки по каждой отдельной карте [29].

На базе проведённого анализа специфики ИНС следует заметить, что это наиболее подходящая для прогнозной аналитики в нефтегазовой компании модель ИАД. Ее ключевыми преимуществами являются:

1. Гибкая возможность конфигурации архитектуры модели и алгоритма обучения на входном наборе статистических данных.
2. Отсутствует необходимость формализации закономерностей в данных и определения зависимостей между ними.
3. Стойкость к наличию шумовых колебаний и влияния на данные, вызывающих противоречия.
4. Оперативность конфигурации модели и ее программного построения с последующим использованием.
5. Автоматизация набора процессов ввода и обработки данных, в том числе нормализации, обучения и тестирования.

б. Функции активации параллелизма посредством использования распределенных вычислений или графических адаптеров [30].

#### Выводы по главе 1

В рамках выполнения данной главы приведены результаты анализа актуальности и проблематики нефтегазовой сферы деятельности, рассмотрены возможности и преимущества использования современных методов интеллектуального анализа данных для решения различных прикладных задач, их логический аппарат и основные этапы. Выполнен анализ особенностей прогнозирования данных, рассмотрены ключевые методы применяемых для различных задач, проведено их сравнение по ряду критериев, обоснована актуальность использования моделей искусственных нейронных сетей и машинного обучения для целевых задач. Проанализированы ключевые положения, методы, модели и функциональные особенности работы моделей нейросетей. На базе проведенных исследований становится возможным дальнейший анализ аналогов, средств разработки и формирование проекта системы, что будет рассмотрено в следующей главе данной работы.

## 2. РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ОТКАЗОВ НАСОСНО-КОМПРЕССОРНЫХ ТРУБ НЕФТЯНЫХ И ГАЗОВЫХ СКВАЖИН

### 2.1 Анализ существующих программных продуктов и систем проведения интеллектуального анализа данных

В настоящее время на рынке программных приложений в области науки о данных для проведения ИАД и построения моделей МО существует ряд функциональных решений. Рассмотрим наиболее популярные и часто используемые на практике системы.

**RapidMiner**, представляет собой кроссплатформенную среду для проведения вычислительных экспериментов и решения задач МО и ИАД [19]. Создаваемые в процессе работы с системой эксперименты описываются в виде расстановки нужного количества соответствующим образом вложенных друг в друга логических операторов посредством использования визуального графического интерфейса и специального окна. Особенности и достоинства системы:

- предоставляет несколько сотен отдельных логических операторов для наиболее популярных методов, а также для импорта, предобработки и визуализации данных;
- интегрирует поддержку операторов библиотеки WEKA;
- реализует встроенный язык написания сценариев, предоставляющий возможности автоматизировать выполнение массивных серии вычислительных экспериментов;
- имплементирует концепцию многоуровневого представления хранимых данных, что позволяет обеспечить эффективную и оперативную работу с поданными источниками данными;
- поддержка подсистемы графического обеспечения визуализации многомерных данных и моделей.

В левой части home-page системы (рис. 2.1) расположены соответствующие блоки компонентов, в частности, панели загрузки данных операторов.

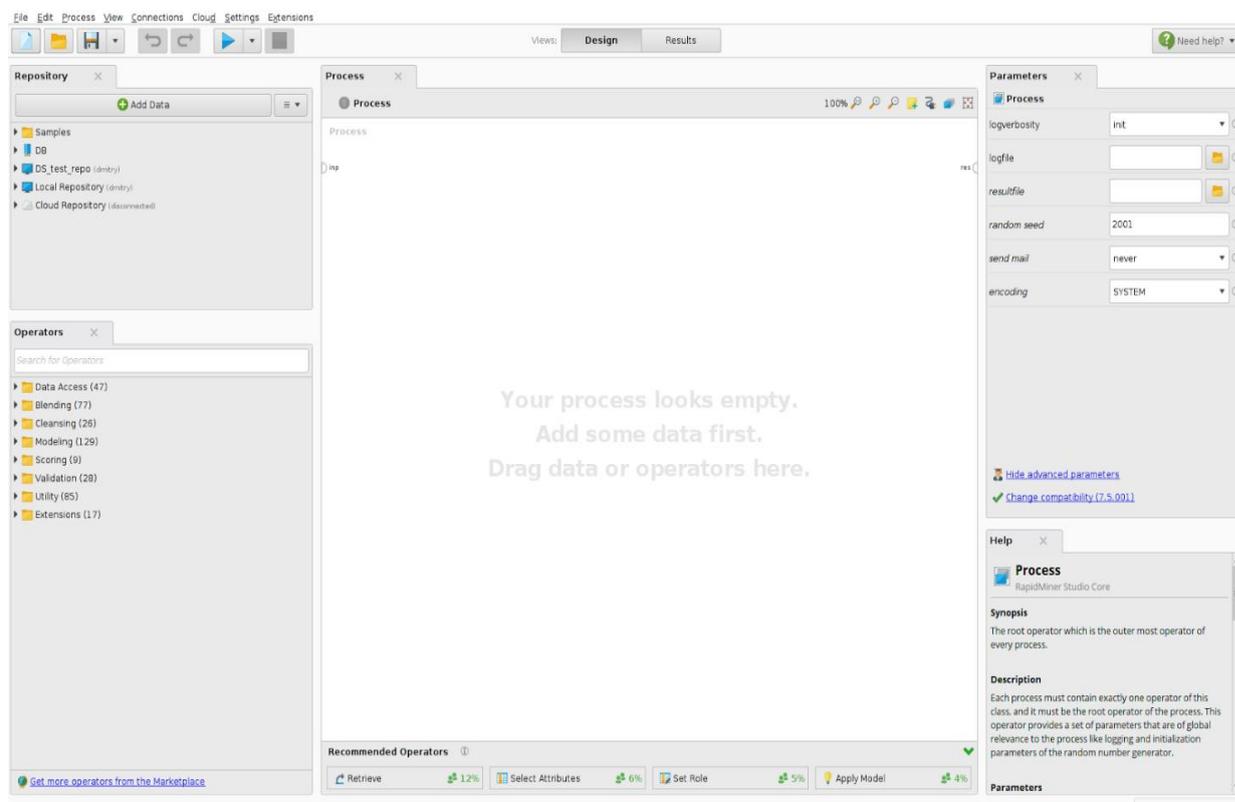


Рисунок 2.1 – Интерфейс системы Rapidminer

RapidMiner реализует возможности удаленной загрузки данных из базы данных или облачных хранилищ (поддерживаются сервисы Amazon, Azure, Dropbox). Структура операторов для удобства пользователя подразделена на такие категории:

- доступ к данным: содержит операторы работа с файлами, базами данных и хранилищами в облаках, а также социальными сетями;
- операторы взаимодействия с атрибутами наборов данных, в частности, для задач преобразования, логических операций и предобработки;
- операторы имплементации различных алгоритмов математического моделирования (оптимизационные, классификационные, модели кластеризации и др.);

- вспомогательные операторы для подключения и выполнения Java и Groovy скриптов, а также для отправки сообщений и планирования событий.

В центральной части рабочего экрана системы расположена прямоугольная рабочая область для размещения компонентов по преобразованию данных и построению цепочки операторов.

С помощью функции drag and drop аналитик может размещать в окне рабочего процесса нужные данные и операторы, которые будут использоваться при построении модели.

Задавая соединяющие линии (связи) между блоками обработки данных и логическими операторами по решению задач ИАД можно формировать последовательность выполнения процесса анализа.

Внизу главного окна расположена панель с подсказками и информированием о результатах проведенных экспериментов. Справа размещена панель, позволяющая осуществлять параметризацию выбранной операции, а также ссылки на документацию указанных параметров и принципов работы системы.

Недостатки системы [20]:

- высокие требования к аппаратным ресурсам;
- не все плагины, расширяющие функционал системы, корректно работают при использовании;
- отсутствие функций по конфигурации параметров для оптимизации скорости работы системы при анализе больших данных.

**Loginom** – является аналитической интеллектуальной платформой, которая предоставляет широкие возможности осуществления глубокой аналитики на базе использования математических моделей и МО для поддержки принятия управленческих решений, основанных на проверенных гипотезах и прогнозах.

Система Loginom является усовершенствованием системы Deductor. Loginom включает в свой состав инструменты для быстрой настройки, подбора и конфигурации параметров анализа, смены логических правил и изменения процессов.

Система поддерживает набор различных технологий, в том числе разведочный статистический анализ, алгоритмы МО, визуализацию результатов. Система Logiном часто используется для краткого статистического анализа [21]. В Logiном встроены возможности, которые выражают ее основные преимущества:

- проведение сложных вычислений по разным формулам;
- консолидация и сведение данных, предобработка и очистка;
- компактность и удобство интерфейса пользователя;
- простота создания и настройки моделей прогнозирования и оптимизации данных;
- интеграция с различными хранилищами данных.

Интерфейс системы Logiном приведен на рис. 2.2.

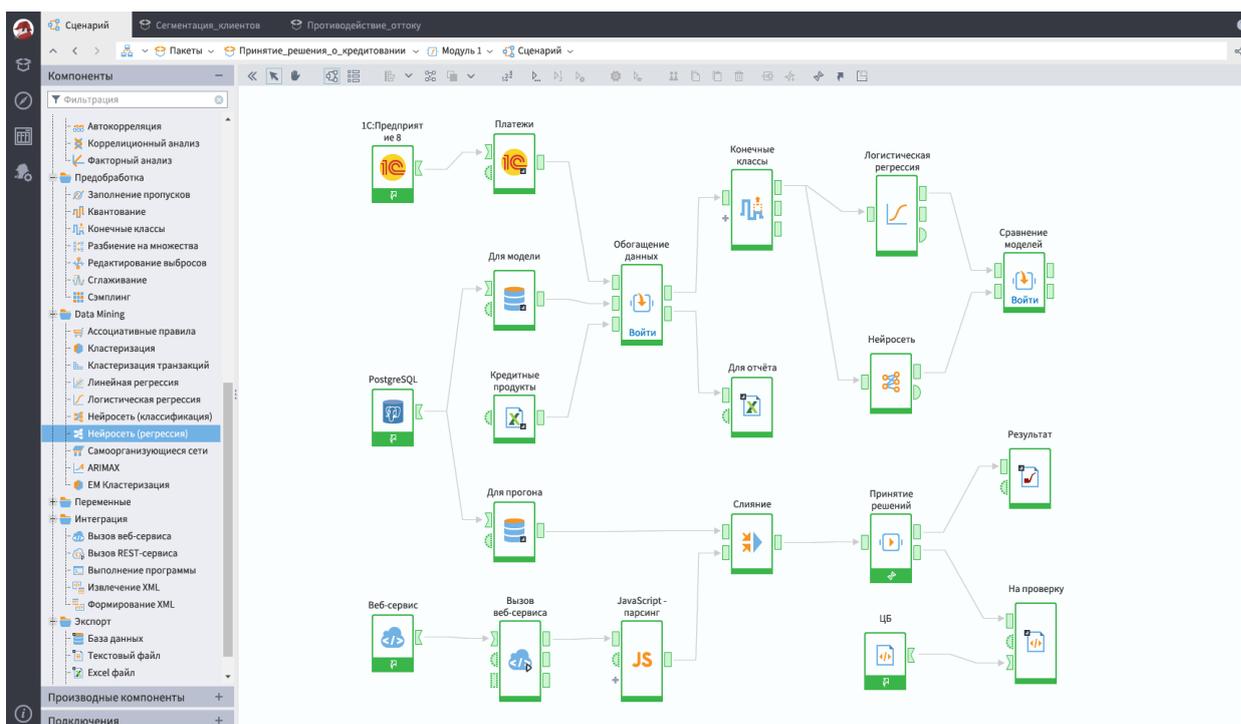


Рисунок 2.2 – Интерфейс системы Logiном

Данная программная платформа имплементирует пользовательский интерфейс, состоящий из наглядных компонентов и инфографических объектов, что позволяет ускорить процесс построения рабочих цепочек анализа данных и проведения глубокой аналитики в сжатые сроки.

Система Logiном по платной подписке предоставляет аналитику набор веб-сервисов и мобильных приложений для взаимодействия пользователя с данными, расширяя возможности построения графических зависимостей для визуализации и интерпретации полученных результатов [22].

Недостатки системы:

- сильно урезанный функционал и множество ограничений по использованию моделей в бесплатной версии;
- отсутствие возможностей подробной конфигурации процессов анализа;
- высокая стоимость корпоративного использования.

**Orange** – это система для визуализации и решения задач ИАД с открытым исходным кодом. ИАД в системе реализуется на основе концепции визуального программирования с поддержкой внедрения программных сценариев на языке Python. Разработанные пользователями системы сценарии могут быть запущены как в терминальном окне, так и в интегрированных средах разработки, в том числе PyCharm и PythonWin. Существует возможность подключения к Jupiter Notebook и инфраструктуре Anaconda. Система включает в свой состав компоненты интерфейса Canvas, где пользователь может размещать виджеты и создавать рабочий процесс анализа данных и построение моделей МО [23]. Интерфейс системы Orange приведен на рис. 2.3. Интегрированные в систему виджеты предлагают пользователю основные функции по обработке, в том числе загрузку и чтение данных, отображение данных в табличном виде, выбор и конфигурация отдельных функций, подключение предикторов обучения, автоматическое сравнение алгоритмов МО по различным метрикам, визуализация элементов обработанных данных и др. Пользователь системы может исследовать визуализации результатов анализа в интерактивном режиме, а также передавать выбранные фрагменты данных или результатов анализа на вход других виджетов [24].

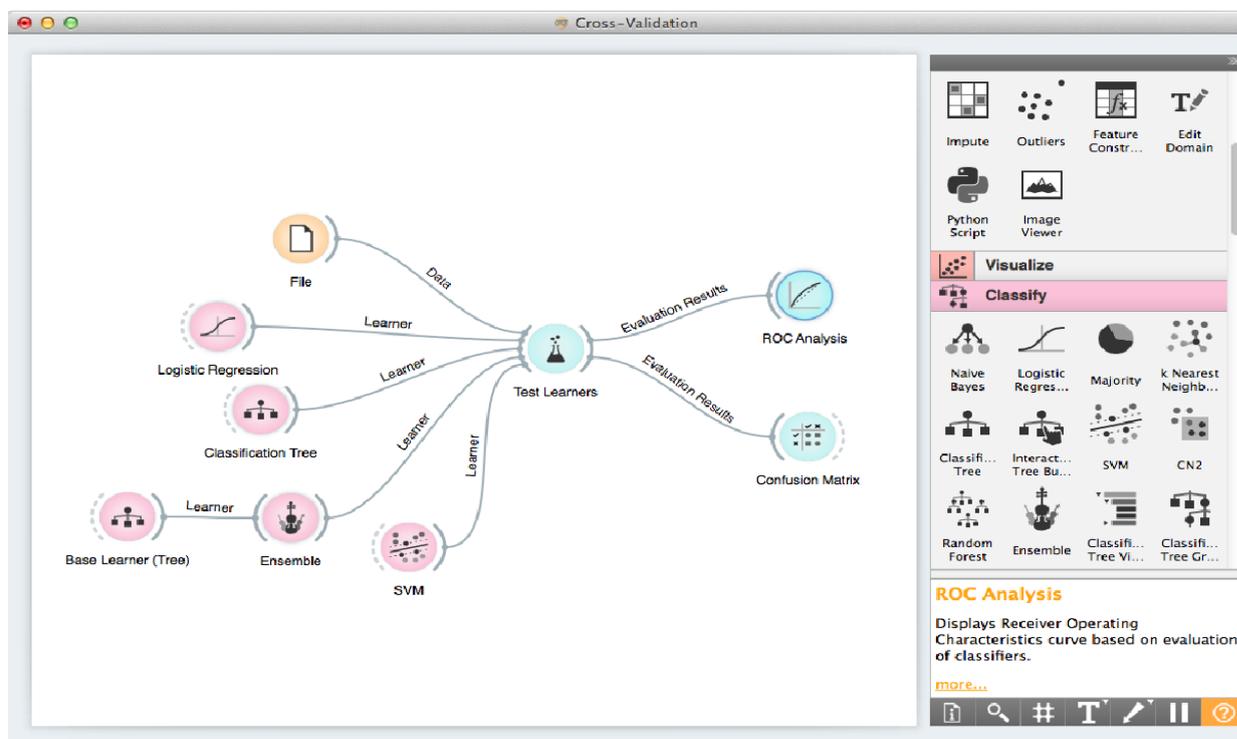


Рисунок 2.3 – Интерфейс системы Orange

В системе Orange, анализ данных выполняется путем объединения компонентов в рабочие процессы. Каждый компонент, называемый виджетом, встраивает некоторую задачу поиска, предварительной обработки, визуализации, моделирования или оценки данных. Например, виджет SQL обращается к данным, хранящимся в базе данных SQL. Он может подключаться к PostgreSQL (требуется модуль psycopg2) или к SQL Server (требуется модуль pumsql).

Особенности системы:

- поддерживается анализ данных с умной визуализацией данных;
- имплементирован традиционный статистический анализ;
- основные алгоритмы МО доступны в интерактивном использовании;
- позволяет автоматизировать процессы формирования представительных выборок;
- поддерживает различные виды кроссвалидации;
- интегрирует популярные метрики оценки качества моделей.

Недостатки системы:

- отсутствуют возможности подробного задания значений для гиперпараметров модели;
- сложности при импорте больших объемов данных;
- невысокая скорость проведения анализа данных;
- отсутствуют функции масштабирования и поддержки распределенных источников.

**Knime Analytics Platform** – представляет собой фреймворк для ИАД, распространяемый по свободной лицензии, однако, имеющий платные дополнения [25]. Данный фреймворк позволяет имплементировать полный жизненный цикл ИАД с использованием методов МО и других утилитарных и операционных функций, в том числе чтение разнородных данных из различных источников и баз данных, преобразование выборок данных и их фильтрацию по различным критериям, алгоритмы анализа, инструменты визуализации и экспорта результатов в различные форматы.

В системе Knime (рис. 2.4) процесс разработки логики процедур анализа данных осуществляется посредством создания Workflow (потока операций), который состоит из набора визуальных компонентов (узлов), выполняющих строго заданную функцию (импорт данных из файла, трансформация, визуализация, сепарация и др.). Для создания связей между узлами их соединяют между собой с помощью стрелок, демонстрирующих направление движение данных в цепочке логических операций.

После создания Workflow модель может быть запущена на исполнение для получения результатов. После этого запуска в базовом сценарии узлы Workflow начинают последовательно запускаться, начиная с самого первого. Если в процессе выполнения отдельного узла произошла ошибка или сбой, то выполнение всех элементов после него прекращается и выдается информационное сообщение. Поддерживаются возможности перезапуска Workflow с указанного узла.

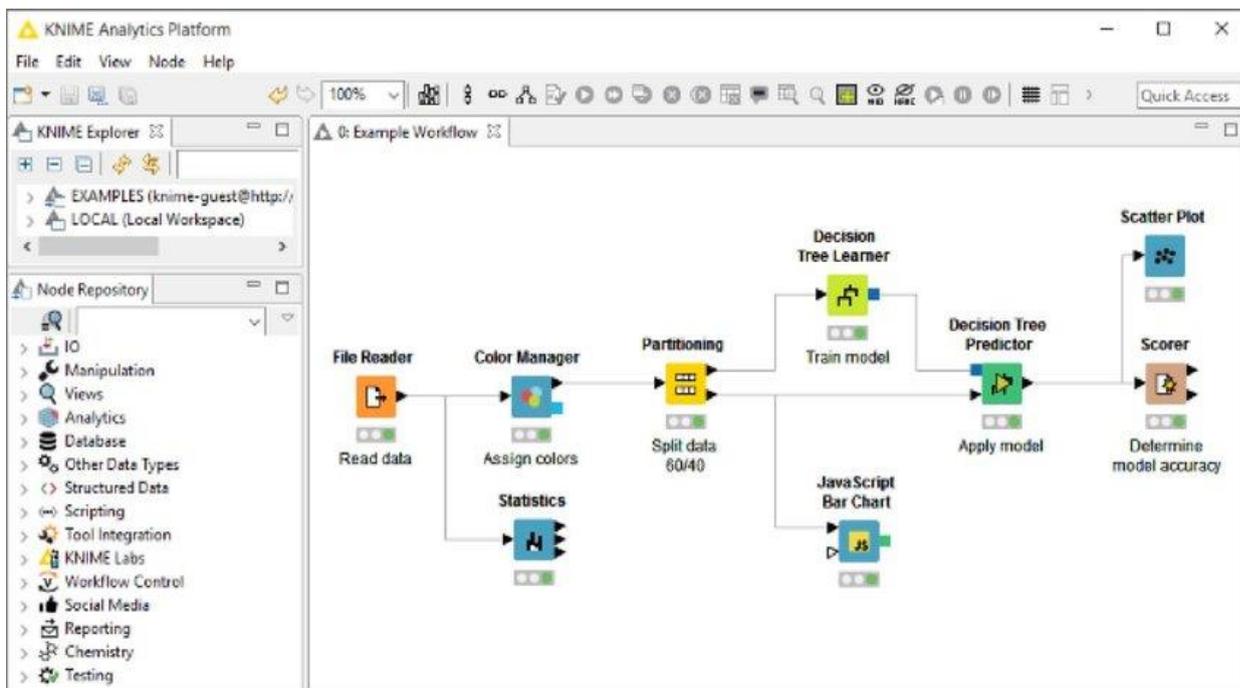


Рисунок 2.4 – Интерфейс системы Kнime

У большинства предусмотренных в системе узлов присутствует опция конфигурационного диалогового окна, где пользователь может выполнить настройку требуемых свойств. Категориями узлов в системе являются: IO – ввод/вывод данных, из файлов или других источников, Manipulation – преобразование и предобработка данных, Views – элементы визуализации данных и результатов анализа, Database – инструменты создания подключений к реляционными БД для чтения или записи данных.

Особенностями системы являются [26]:

- размещение и подключение Workflow на сервере с передачей прав доступа к полученным результатам анализа посредством REST API. Эта опция доступна в случае платной подписки, в частности, после покупки отдельного KNIME-Server;
- интеграция большого числа сторонних библиотек, которые становятся доступны пользователю в виде отдельных узлов;
- поддержка выполнения операции Pivot напрямую в БД или на данных локального кэша;
- большая библиотека примеров;

- инструменты использования Hadoop.

Недостатки системы:

- недостаточно гибкое поведение узлов, отсутствуют возможности оперативной и тонкой конфигурации ряда параметров;
- перегрузка ряда окон элементами интерфейса, что увеличивает время анализа.

Рассмотренные системы позволяют подсказывать следующее действие в цепочке подготовки данных для анализа, тренировки модели МО, её валидации и оценки точности. Преимуществами данных систем является поддержка автоматического исправления некоторых ошибок в процессе создания моделей.

Однако, рассмотренные системы достаточно сложны в использовании, не позволяют обеспечить полноценную степень интерактивности процесса построения моделей МО, а также являются достаточно сложными при установке и развертывании, а также ресурсоемкими. В связи с этим, актуальной задачей является проектирование и разработка собственного программного обеспечения на базе современных средств программирования.

## 2.2 Анализ ключевых преимуществ программных библиотек поддержки процессов анализа данных

Разработка программ представляет собой довольно сложный процесс, главной целью которого является создание и дальнейшее совершенствование (сопровождение) программного кода, обладающего высоким уровнем надежности и качества. Средства разработки ПО представляют собой некоторую совокупность приемов, методик, подходов, методов, а также инструментальных программных приложений (компиляторы, библиотеки, модули), используемые разработчиком ПО для написания функционального кода программы, удовлетворяющего поставленным требованиям.

Для проведения исследований в рамках данной работы выбран язык программирования Python. Python является объектно-ориентированным

высокоуровневым языком программирования с поддержкой динамической типизации, имплементацией процесса автоматического управления памятью, различных структур данных (кортежи, словари и списки) и активно развивается.

NumPy – библиотека, которая позволяет эффективно обрабатывать многомерные массивы. NumPy предназначена для соединения возможностей организации высокой скорости вычислений (как это можно увидеть в языках C и Fortran) со скоростью разработки прикладных программ на языке Python. Принципиальным достоинством NumPy является составленный набор готовых математических функций, которые вмещают в себя ряд ключевых разделов математики. Большая часть таких функций может быть оптимизирована под нужную задачу. Сам механизм трансляции NumPy организует возможности написания компактного кода и незначительного количества вложенных циклов.

Pandas – это высокоуровневая Python библиотека для анализа данных. В экосистеме Python, pandas является наиболее продвинутой и быстроразвивающейся библиотекой для обработки и анализа данных. Чтобы эффективно работать с pandas, необходимо освоить самые главные структуры данных библиотеки: DataFrame и Series. Структура/объект Series Pandas, представляет собой высокоуровневую библиотеку для разведывательного анализа и обработки данных из различных файлов с поддержкой табличных форматов представления.

SciKit-Learn – представляет собой некоторый набор программных зависимостей и библиотек с функционалом по решению прикладных МО задач. Благодаря тому, что библиотека создана средствами Python, ее интеграция в проекте является быстрой и простой, дополнительным преимуществом является оптимизация выполнения ряда алгоритмов, благодаря использованию вставок кода на Cython (написание модулей C/C++ для Python). Удобством библиотеки являются обширные объемы примеров использования ее возможностей, статей и документации на официальном сайте. Дополнительным преимуществом является поддержка программных модулей быстрой визуализации данных, в том числе для оценки качества результатов решения задач МО [43].

Matplotlib – это библиотека на языке Python для визуализации данных. В ней можно построить двумерные (плоские) и трехмерные графики. Matplotlib – альтернатива модулю визуализации программы для технических вычислений MatLab. У Matplotlib объектно-ориентированный интерфейс, то есть, пользователь напрямую взаимодействует с каждым объектом. С помощью кода можно задавать любой элемент диаграммы, в том числе ярлыки и отметки на осях. Библиотека содержит большой набор инструментов для двумерной графики. Модуль `matplotlib.pyplot` предоставляет процедурный интерфейс к (объектно-ориентированной) библиотеке `matplotlib`, который во многом копирует инструменты пакета MATLAB. Инструменты модуля `pyplot`, де-факто, являются стандартным способом работы с библиотекой `matplotlib` [29].

Tensorflow (TF) – представляет собой современных и функциональный набор зависимостей для эффективного решения задач глубокого МО от компании Google, распространяющийся на бесплатной основе. Весь процесс взаимодействия с объектами TF основан на процесс формирования и построения графов вычислений для формализации вычислительных задач над массивами данных [44]. Основа TF формируется на базе инициализации общей структуры для проведения последовательных вычислительных операций. Разрабатываемые на основе TF структуры подразделяются на потоки формирования графов вычислений и проведения вычислительных расчётов в выделенных для этого структурах данных. Каждый отдельный граф включает в себя так называемые «плейсхолдеры» в качестве основных элементов, необходимых для построения моделей графов с манипуляцией тензорами (многомерными массивами, необходимыми для представления графов и оперирования их содержимым).

Keras – это функциональная библиотека, реализующая высокоуровневый API для создания объектов ИНС. Преимуществом данного решения является быстрое действие и низкая связность зависимостей внутри библиотеки, а также кроссплатформенность и поддержка разных версий языка Python. Таким образом, проведя анализ рассмотренных средств, следует отметить, что наиболее подходящей для задач прогнозной аналитики в нефтегазовой компании является

SciKit-Learn [45], так как

- библиотека обладает функциональных и развернутым API;
- реализует поддержку имплементации всех стадий решения задач классификации;
- поддерживает возможности разнообразного анализа и обработки данных с последующей интерактивной визуализацией результатов;
- эффективно реализует парадигмы горизонтального масштабирования на серверы кластеров.

Дополнительным удобным средством разработки на языке Python является Jupyter Notebook, который представляет собой унифицированное веб-приложение, поставляющееся на рынок по открытой лицензии, позволяющее создавать интерактивные документы, включающие в свой состав программный код на языке Python, расчетные формулы, графические визуализации, диаграммы и справочное текстовое описание. Дополнительным преимуществом является поддержка возможностей обмена созданными документами в формате `.ipynb` и их открытие на разных платформах [29].

Использование данного приложения включает в состав процессы очистки, преобразования входных данных, проведение численного и статистического моделирования, визуализацию данных и решение задач МО. Jupyter Notebook поддерживает интеграцию других ЯП кроме Python, что обеспечивает расширение возможностей ядра, а также дополнительные команды [30]. Интерфейс среды Jupiter Notebook приведен на рис.2.7.

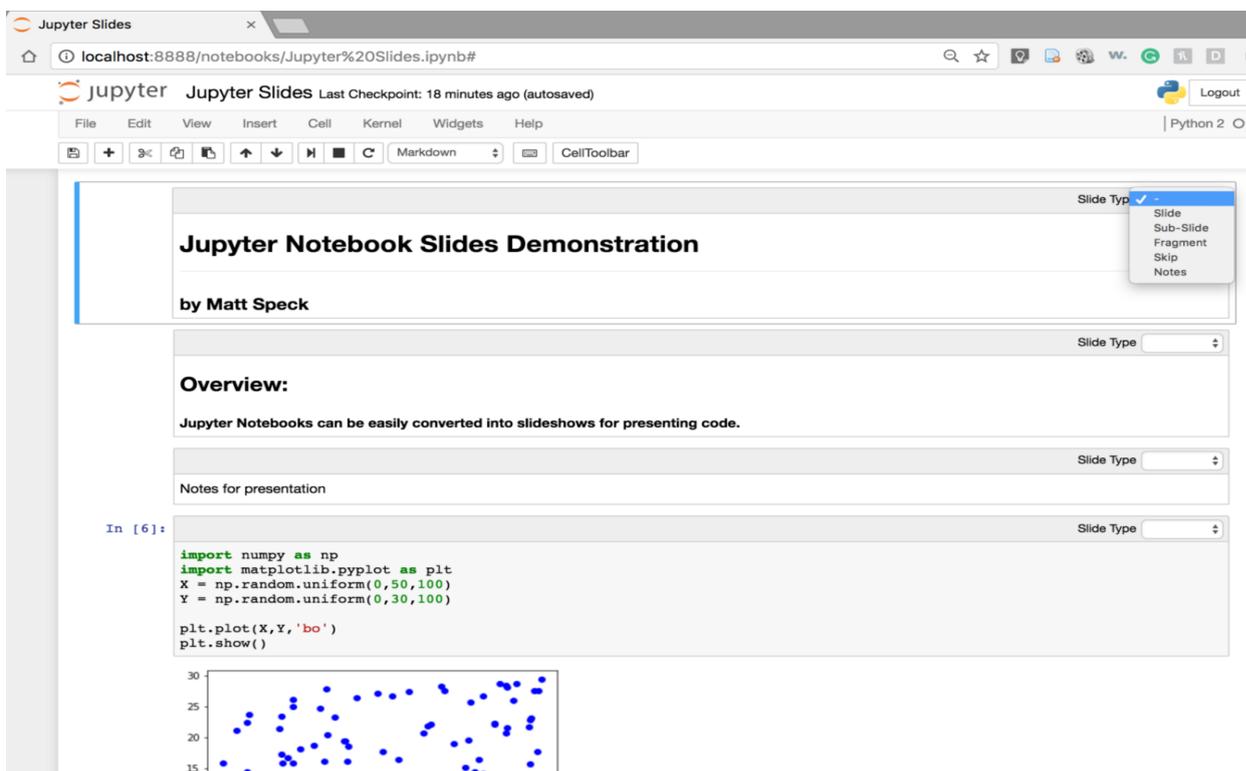


Рисунок 2.7– Интерфейс среды Jupyter Notebook

Jupyter Notebook реализует дополнительные, по сравнению со стандартной IDLE Python средой, возможности по созданию проекта, заключающиеся в следующем [31]:

- поддержка интерактивной оболочки;
- имплементирует основное ядро и ряд более гибких инструментов интерфейса Project Jupyter;
- поддерживает функции автоматического анализа объекта (проверка свойства объекта во время выполнения);
- подсветка синтаксиса кода;
- архивация истории взаимодействий и изменений;
- поддержка заполнения вкладок ключевыми словами, автозаполнение имен переменных и методов;
- поддержка возможностей работы в многопоточном режиме;
- интеграция в другие программы на языке Python;
- интерактивные функции отладки кода Python.

Jupyter Notebook поддерживает функции отображения и редактирования ряда форматов данных, в том числе изображений, текстовых файлов передачи данных CSV, JSON и др. Приложение может быть дополнено функционалом с помощью интеграции и установки функциональных расширений, реализующих настройку различных компонентов под свои нужды, в том числе темы, редакторы файлов и дополнительные функции экспорта.

### 2.3 Построение UML-диаграмм для отображения статической структуры системы, поведенческих и физических аспектов системы

Для моделирования основных возможностей системы применим язык UML. Для этого воспользуемся онлайн-сервисом VisualParadigm.

Диаграмма вариантов использования проектируемой системы приведена на рис.2.8.

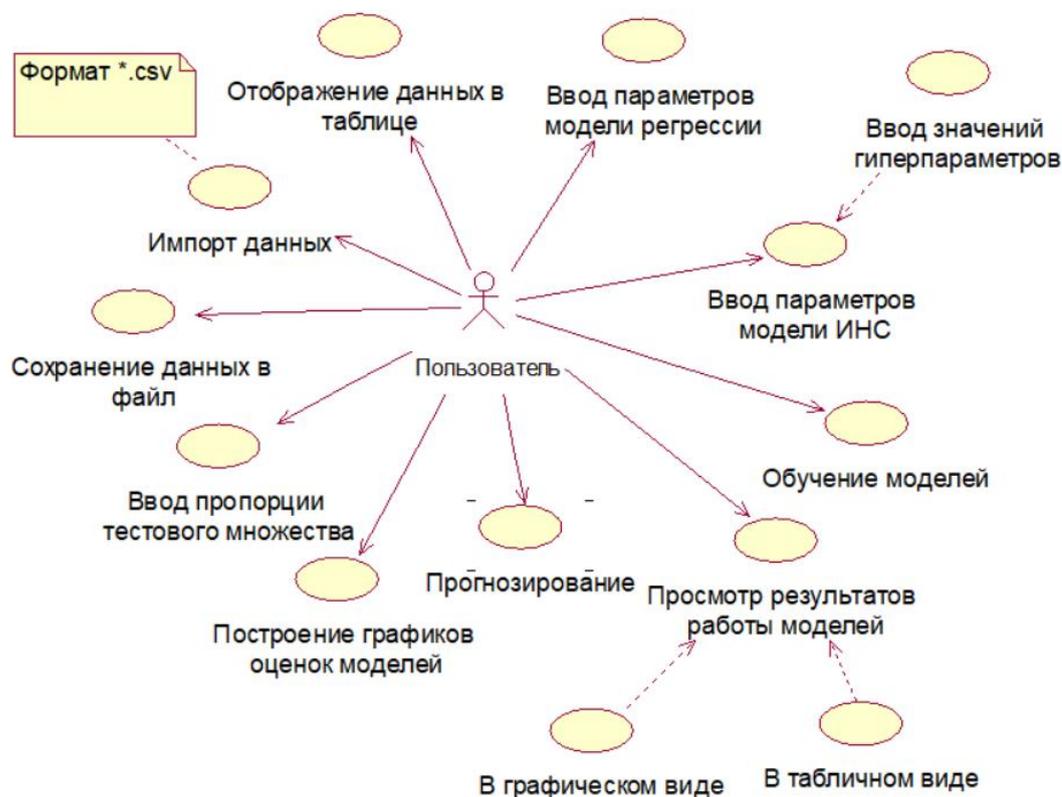


Рисунок 2.8– Диаграмма вариантов использования системы

Пользователь имеет возможности импорта данных, просмотра данных в табличном виде, ввод пропорции тестового множества, ввода значений гиперпараметров моделей регрессии и ИНС, запуска процесса обучения моделей, просмотр результатов работы моделей, построения и просмотра графиков оценок процесса обучения моделей и сохранения их в виде графических файлов; осуществление ввода данных для построения прогноза.

Для моделирования взаимодействия между логическими модулями программного обеспечения и обозначения процессов передачи и обмена данными между ними целесообразно применение диаграмм последовательности действий. Диаграмма последовательностей действий обучения и использования модели МО приведена на рис.2.9.

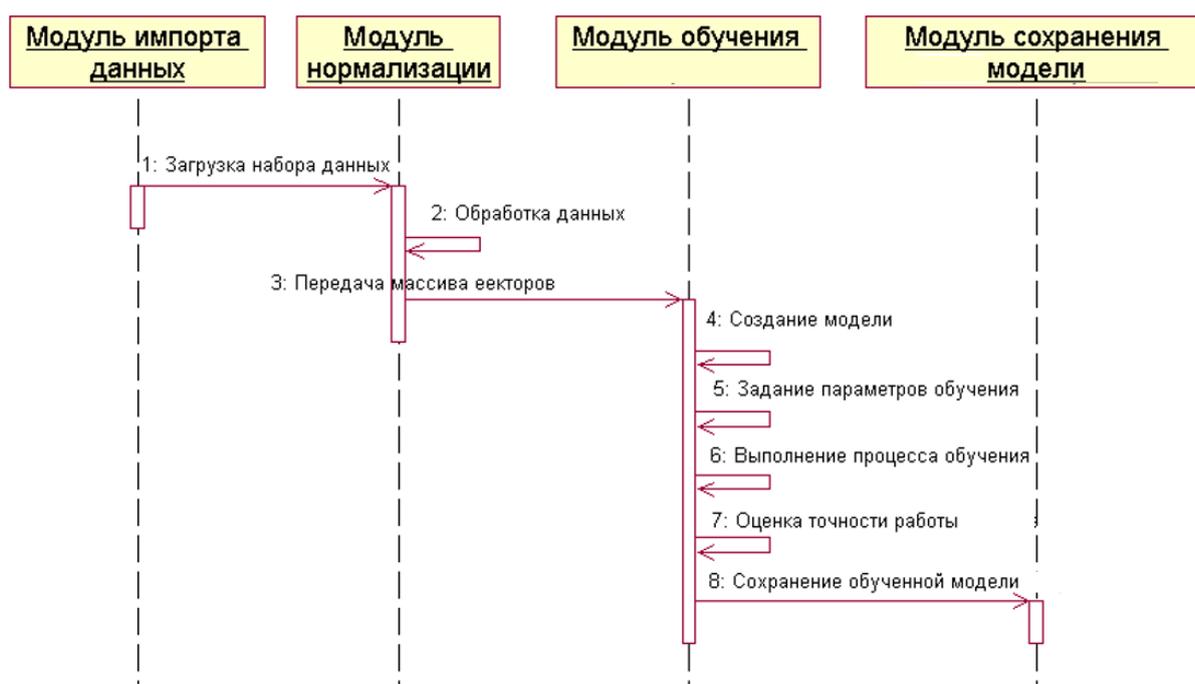


Рисунок 2.9 – Диаграмма последовательности действий обучения модели МО

Поступающий через модуль импорта данных dataset загружается в программу и с помощью модуля нормализации обрабатывается для создания итогового массива векторов для всех признаков набора.

Переданные данные на вход модулю обучения моделей МО используются в качестве основы для ее обучения после создания и конфигурации. На основании

проведения процесса обучения осуществляется оценка точности работы модели МО и построение графика оценки качества ее работы.

После обучения и сохранения модели МО становится возможным ее использование для прогнозирования значений выходной переменной и отображения полученных результатов с помощью модуля интерпретации.

Диаграмма компонентов программного приложения приведена на рис. 2.10. Главное приложение использует ряд внешних компонентов (зависимостей): `scipy` – для проведения процессов обработки данных; `sklearn`, `NumPy`, `Pandas` – для построения структур данных и алгоритмов МО; `matplotlib` и `seaborn` – для визуализации результатов оценки качества прогнозирования, `Keras` и `Tensorflow` – для реализации логики построения, обучения и тестирования объектов моделей ИНС и нормализации данных соответственно.

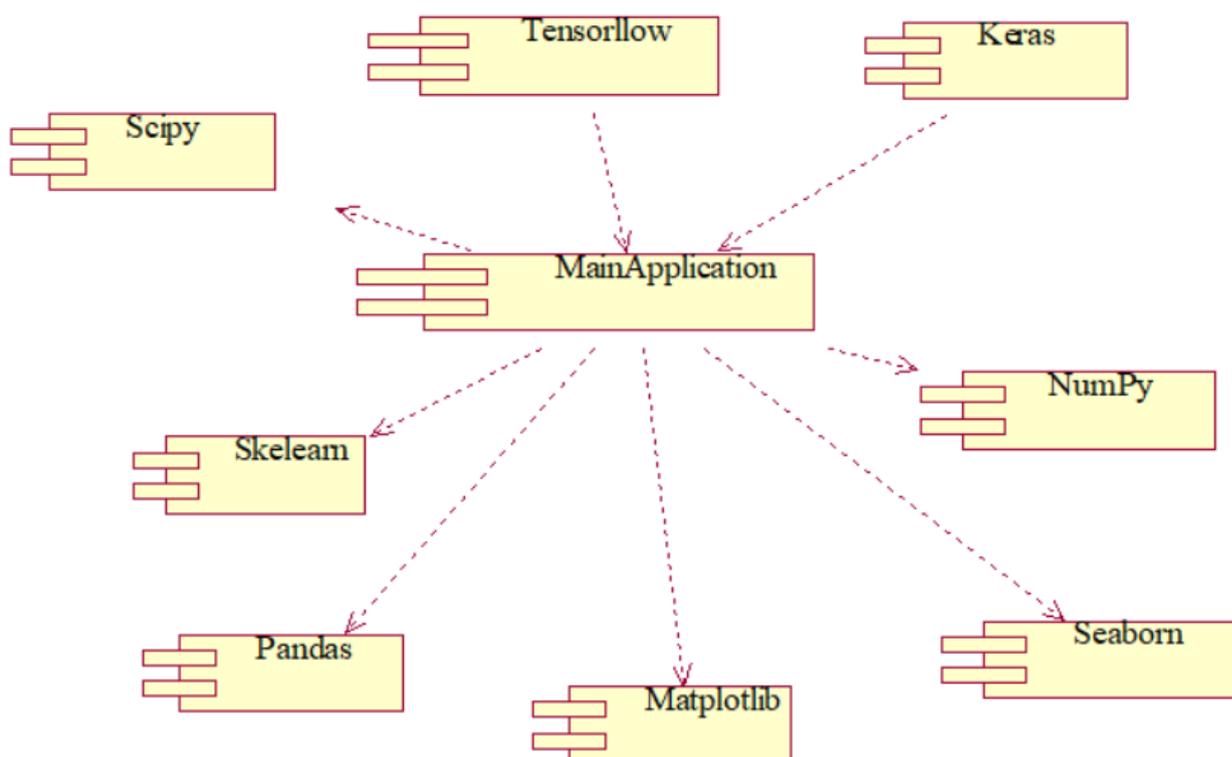


Рисунок 2.10 – Диаграмма ключевых компонентов системы

Диаграмма развертывания приложения приведена на рис.2.11.

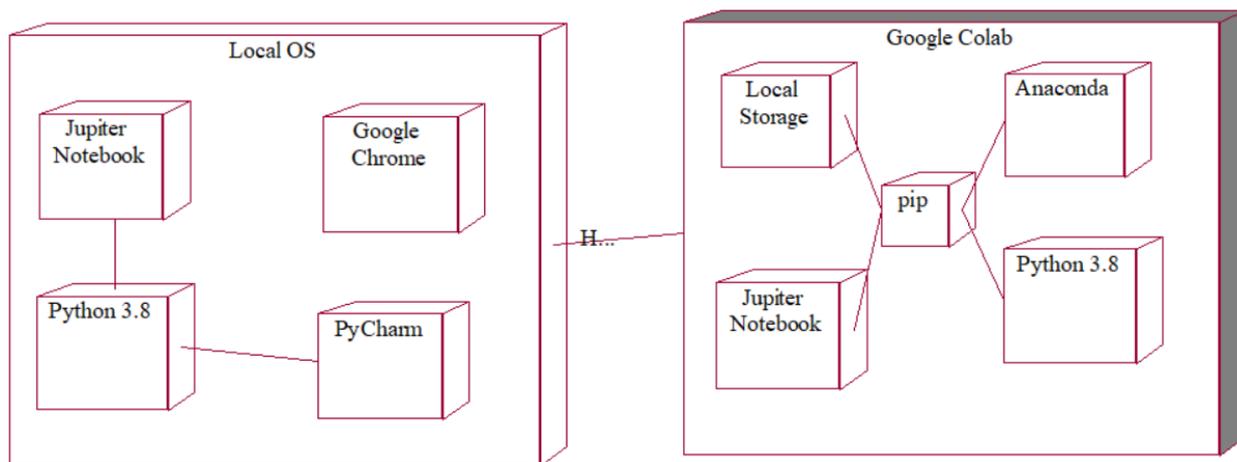


Рисунок 2.11 – Диаграмма развертывания приложения

Локально разработка производится посредством использования языка программирования Python версии 3.8 и среды разработки PyCharm. Посредством использования браузера Google Chrome при авторизации через учетную запись gmail производится вход в облачный веб-сервис Google Colab для развертывания, тестирования и исследования работы программного приложения в виде файла .ipynb. При создании модели ИНС на вход системе подаются входная выборка и набор параметров, на выходе формируется модель нейросети и набор весовых коэффициентов связей между нейронами.

Для обучения созданной модели ИНС используется алгоритм обратного распространения ошибки. Алгоритм является одним из методов обучения многослойных нейронных сетей прямого распространения. Обучение алгоритмом обратного распространения ошибки предполагает два прохода по всем слоям сети: прямого и обратного. При прямом проходе входной вектор подается на входной слой модели нейронной сети, после чего распространяется по сети от слоя к слою. В результате генерируется набор выходных сигналов, который и является фактической реакцией сети на данный входной образ. Во время прямого прохода все синаптические веса сети фиксированы. Во время обратного прохода все синаптические веса настраиваются в соответствии с правилом коррекции ошибок, а именно: фактический выход сети вычитается из желаемого, в результате чего формируется сигнал ошибки. Этот сигнал, в последствии, распространяется по сети в направлении, обратном направлению синаптических связей.

Для моделирования взаимодействия между логическими модулями системы и обозначения процессов передачи и обмена данными между ними целесообразно применение диаграмм последовательности действий. Разработанная диаграмма последовательности действий обучения модели ИНС приведена на рис.2.12.

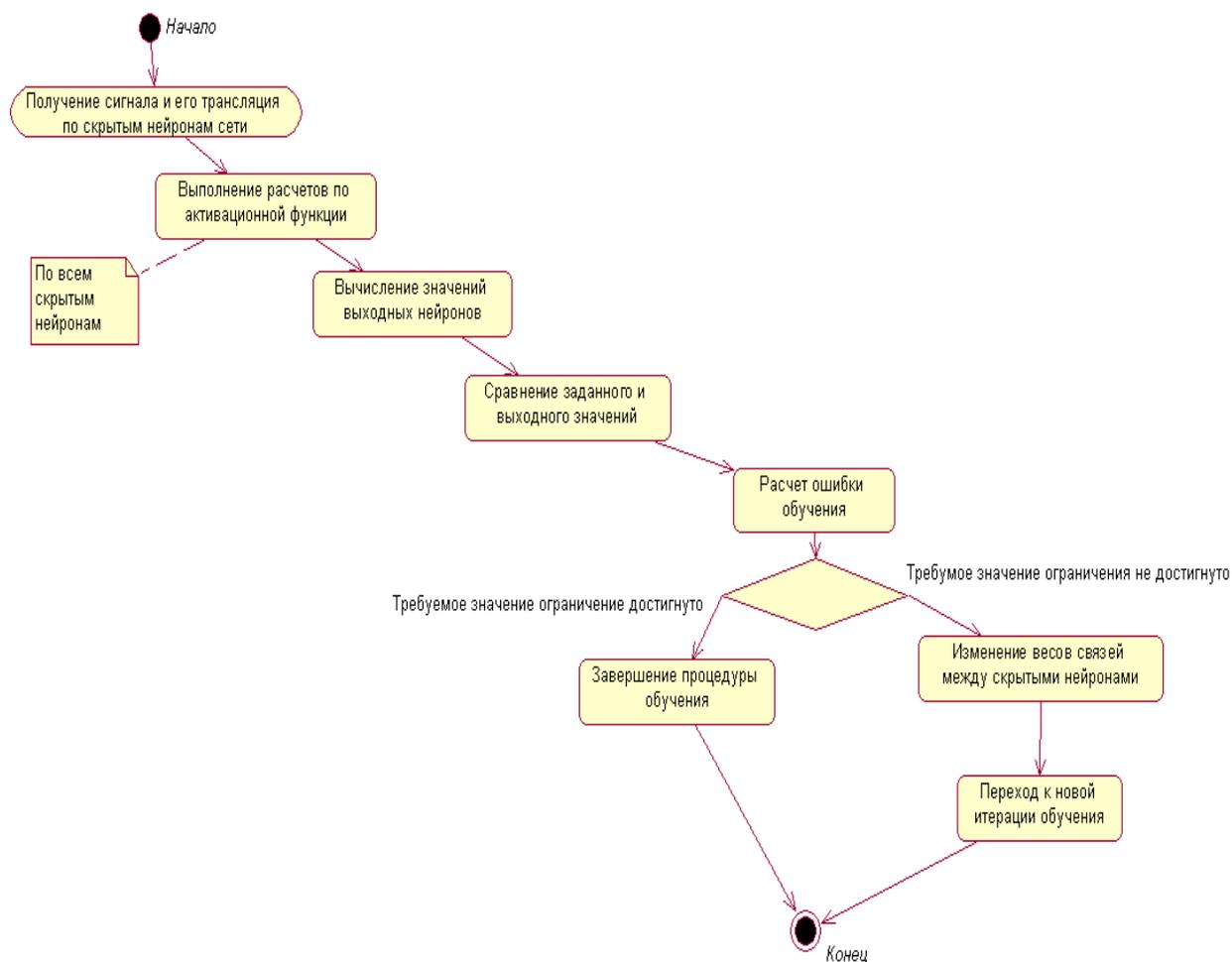


Рисунок 2.12 – Диаграмма активностей алгоритма обучения ИНС

Порядок выполнения алгоритма обратного распространения ошибки в общем виде имеет вид:

- Инициализация синаптических весов случайными значениями.
- Выбор обучающей пары из обучающего множества и передача входного вектора данных на вход сети.
- Вычислить значение на выходных нейронах сети.
- Вычислить отличия между выходными значениями сети и требуемым

значением.

- Модификация весов сети для минимизации ошибки.
- Повторять указанные шаги для каждого вектора обучающего множества до тех пор, пока ошибка на всем множестве не достигнет приемлемого уровня.

Обобщенный алгоритм нормализации данных при обучении моделей ИНС приведен на рис.2.13.



Рисунок 2.13 – Обобщенный алгоритм нормализации данных

Поступающая выборка данных обрабатывается на этапе получения входных данных, после чего обработанный и нормализованный набор данных в бинарном виде поступает на вход процесса инициализации модели ИНС, на выходе которой формируется обученная модель. Затем осуществляется оценка точности ее работы. Далее, в случае, если значения оценок точности работы модели являются высокими, обученная модель может быть сериализована.

## 2.4 Обобщённая функциональная структура системы прогнозирования

На основании проведенного анализа и определения требований к программному обеспечению разработана обобщённая функциональная структура интеллектуальной системы прогнозирования отказов насосно-компрессорных труб нефтяных и газовых скважин (рис. 2.14).

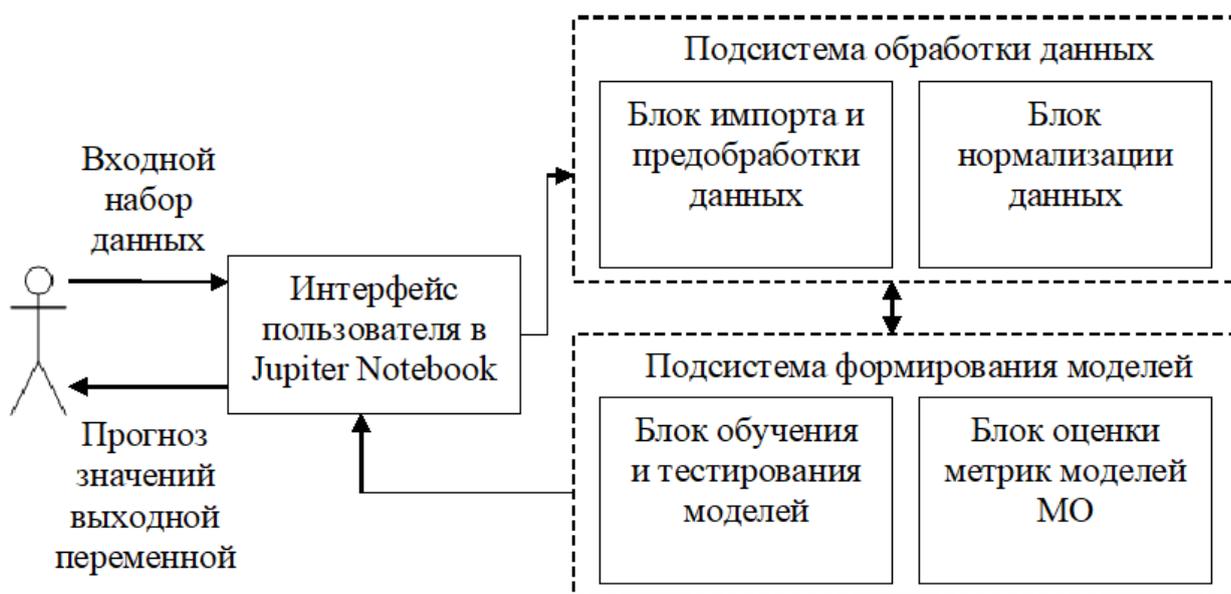


Рисунок 2.14 – Функциональная структура системы прогнозирования

Система включает следующие компоненты:

1. Подсистема обработки данных:
  - а) блок импорта и предобработки данных;
  - б) блок нормализации данных;
2. Подсистема формирования моделей МО:
  - а) блок обучения и тестирования моделей;
  - б) блок оценки и визуализации метрик моделей.

Рассмотренные подсистемы и блоки имплементируют в себя модульную структуру системы. Схема взаимосвязи ключевых модулей системы приведена на рис.2.15.

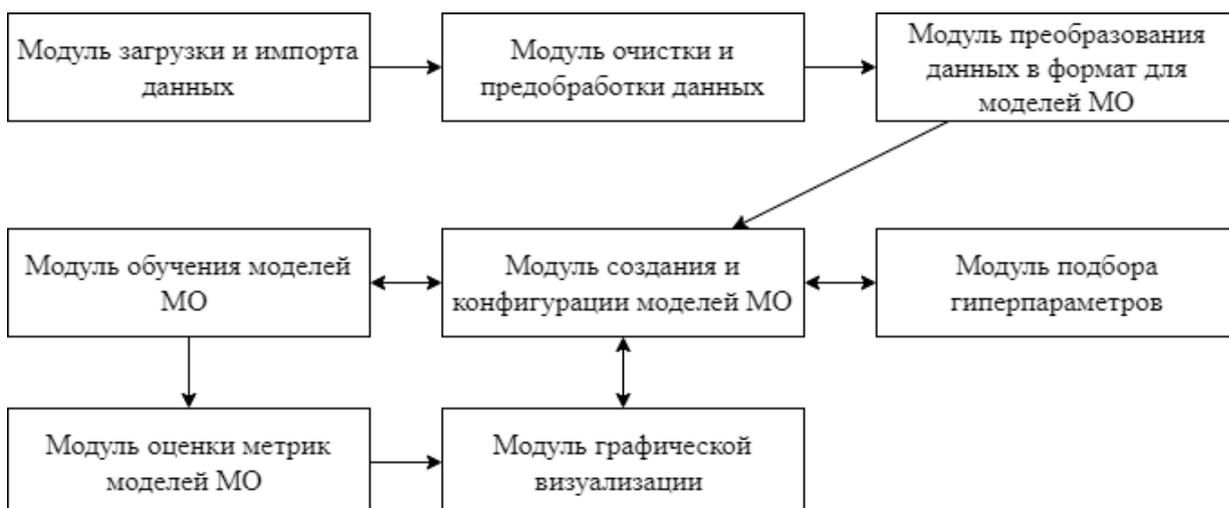


Рисунок 2.15 – Структурная схема модулей системы

Разрабатываемая система должна обеспечивать выполнение задач, приведенных в таблице 2.1.

Таблица 2.1 – Задачи системы

Задача	Действие
Ввод данных	<ul style="list-style-type: none"> <li>– Разработчик указывает путь к файлу набора данных.</li> <li>– ПО осуществляет открытие файла в виде отдельного потока в режиме чтения.</li> <li>– ПО выполняет импорт данных в используемые динамические структуры данных для их хранения и обработки в программе.</li> <li>– ПО завершает работу потока чтения и закрывает файл.</li> </ul>

<p>Нормализация данных</p>	<ul style="list-style-type: none"> <li>– ПО осуществляет создание новой структуры данных объекта pandas dataframe.</li> <li>– ПО выполняет приведение категориальных и номинальных признаков к числовому виду.</li> <li>– ПО обновляет содержимое объекта итоговой коллекции данных.</li> </ul>
<p>Создание моделей МО</p>	<ul style="list-style-type: none"> <li>– Разработчик создает объект выбранного типа МО.</li> <li>– Разработчик осуществляет конфигурацию модели МО путем задания значений и числа гиперпараметров.</li> <li>– ПО сохраняет структуру модели в виде отдельного объекта.</li> </ul>
<p>Обучение созданной модели</p>	<ul style="list-style-type: none"> <li>– Разработчик выбирает и конфигурирует алгоритм обучения модели МО.</li> <li>– Разработчик указывает нормализованный набор данных для обучения и подает его на вход модели.</li> <li>– Разработчик задает ограничения на остановку обучения.</li> <li>– ПО осуществляет обучение модели на поданной входной выборке.</li> <li>– ПО выполняет проверку достижения заданных ограничений и, в случае успеха, сохраняет обученную модель в виде объекта.</li> </ul>
<p>Тестирование созданной модели</p>	<ul style="list-style-type: none"> <li>– Разработчик указывает переменную, хранящую объект обученной модели МО.</li> <li>– Разработчик указывает тестовую выборку для проведения валидации модели.</li> <li>– ПО проводит тестирование выбранной модели на тестовой выборке, фиксируя значения метрик оценки ее точности.</li> </ul>

<p>Визуализация результатов работы модели</p>	<ul style="list-style-type: none"> <li>– Разработчик указывает переменную, содержащую результаты оценки точности моделей.</li> <li>– Разработчик подключает объекты библиотеки визуализации данных.</li> <li>– ПО осуществляет построение графических диаграмм.</li> </ul>
---	--

## Выводы по главе 2

Приведено описание разработки интеллектуальной информационной системы прогнозирования отказов насосно-компрессорных труб нефтяных и газовых скважин как средства повышения новых подходов к экономии материальных затрат и продления их работоспособности. Выполнен анализ существующих программных продуктов и систем проведения интеллектуального анализа данных, проанализированы возможности использования языков программирования для решения задач анализа данных, осуществлен анализ ключевых преимуществ программных библиотек поддержки процессов анализа данных. В качестве основы для проведения исследования и вычислительных экспериментов выполнено построение UML-диаграмм для отображения статической структуры системы, поведенческих и физических аспектов системы, а также построена ее обобщённая функциональная структура, описаны ключевые модули и подсистемы. На базе полученных результатов становится возможным описание используемого набора данных и проведение исследования работы моделей МО для решения рассматриваемой задачи.

## **4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение**

### **Введение**

Анализ данных для прогнозирования частоты отказов насосно-компрессорных труб позволяет сократить количество аварий на скважинах, а также снизить количество случаев подъема из скважины трубы, годной к эксплуатации, что уменьшит затраты нефтегазовой компании как на закупку новой трубы, так и на проведение ремонтов скважин.

Целью данного раздела является определение перспективности исследовательского проекта.

Данный раздел, предусматривает рассмотрение следующих задач:

- Оценка коммерческого потенциала разработки.
- Планирование научно-исследовательской работы;
- Расчет бюджета научно-исследовательской работы;
- Определение ресурсной, финансовой, бюджетной эффективности исследования.

Цель данной ВКР – исследование методов анализа данных в нефтегазовой компании на основе методов машинного обучения для прогнозирования частоты отказов насосно-компрессорных труб.

### **4.1 Оценка коммерческого потенциала и перспективности проведения исследований с позиции ресурсоэффективности и ресурсосбережения**

#### **4.1.1 Анализ конкурентных технических решений**

В ходе исследования был рассмотрен вариант замены насосно-компрессорных труб на основе графиков планово-предупредительных ремонтов.

Детальный анализ необходим, т.к. каждый тип покрытия имеет свои достоинства и недостатки. В таблице 4.1 показано сравнение разработок-конкурентов и разработки данного НИ с точки зрения технических и экономических критериев оценки эффективности.

Таблица 4.1 – Сравнение конкурентных технических решений (разработок)

Критерии оценки	Вес критерия	Баллы		Конкурентоспособность	
		Б <sub>Ф</sub>	Б <sub>К1</sub>	К <sub>Ф</sub>	К <sub>К1</sub>
1	2	3	4	5	6
<b>Технические критерии оценки ресурсоэффективности</b>					
Актуальность	0,2	5	4	1	0,8
Удобство в эксплуатации	0,1	5	2	0,5	0,2
Простота разработки	0,1	2	5	0,2	0,5
Эффективность работы	0,1	5	3	0,5	0,3
Надежность продукта	0,1	4	2	0,4	0,2
<b>Экономические критерии оценки эффективности</b>					
Стоимость разработки	0,2	4	5	0,8	1
Предполагаемый экономический эффект	0,2	5	3	1	0,6
<b>Итого</b>	<b>1</b>	<b>30</b>	<b>24</b>	<b>4,4</b>	<b>3,6</b>

Расчет конкурентоспособности, на примере стабильности срабатывания, определяется по формуле:

$$K = \sum B_i \cdot B_j = 4,4,$$

где  $K$  – конкурентоспособность проекта;  $B_i$  – вес показателя (в долях единицы);  $B_j$  – балл показателя.

Проведенный анализ конкурентных технических решений показал, что исследование является наиболее актуальным и перспективным, имеет конкурентоспособность.

#### 4.1.2 SWOT-анализ

Для исследования внешней и внутренней среды проекта, в этой работе проведен SWOT-анализ с детальной оценкой сильных и слабых сторон исследовательского проекта, а также его возможностей и угроз.

Первый этап, составляется матрица SWOT, в которую описаны слабые и сильные стороны проекта и выявленные возможности и угрозы для реализации проекта, которые проявились или могут появиться в его внешней среде, приведены в таблице 4.2.

Таблица 4.2 – Матрица SWOT-анализа

<b>Сильные стороны</b>	<b>Слабые стороны</b>
С1. Улучшение показателей.	Сл1. При некачественных исходных данных возможны погрешности при анализе.
С2. Не требуется специализированного оборудования.	Сл2. Ограниченный функционал конечного ПО.
С3. Более свежая информация, которая была использована для разработки проекта.	Сл3. Малый опыт создания подобных систем.
С4. Невысокие системные требования.	
<b>Возможности</b>	<b>Угрозы</b>
В1. Невысокий уровень конкуренции.	У1. Снижение стоимости разработок конкурентов.
В2. Появление потенциального спроса на новые разработки.	У2. Появление зарубежных аналогов и более ранний их выход на рынок.

На втором этапе на основании матрицы SWOT строятся интерактивные матрицы возможностей и угроз, позволяющие оценить эффективность проекта, а также надежность его реализации. Соотношения параметров представлены в таблицах 4.3–4.6.

Таблица 4.3 – Интерактивная матрица проекта «Возможности проекта и сильные стороны»

<b>Возможности проекта</b>		C1	C2	C3	C4
	B1	-	-	-	-
	B2	+	-	+	-

Таблица 4.4 – Интерактивная матрица проекта «Возможности проекта и слабые стороны»

<b>Возможности проекта</b>		Сл1	Сл2	Сл3
	B1	-	-	+
	B2	-	-	-

Таблица 4.5 – Интерактивная матрица проекта «Угрозы проекта и сильные стороны»

<b>Угрозы проекта</b>		C1	C2	C3	C4
	У1	-	+	-	+
	У2	-	+	-	+

Таблица 4.6 – Интерактивная матрица проекта «Угрозы проекта и слабые стороны»

<b>Угрозы проекта</b>		Сл1	Сл2	Сл3
	У1	+	-	-
	У2	-	-	-

Результаты анализа представлены в итоговую таблицу 4.7.

Таблица 4.7 – Итоговая таблица SWOT-анализа

	<b>Сильные стороны научно-исследовательского проекта</b> С1. Улучшение показателей.	<b>Слабые стороны научно-исследовательского проекта</b> Сл1. При некачественных исходных данных возможны погрешности при анализе.
--	--	--

	<p>С2. Не требуется специализированного оборудования.</p> <p>С3. Более свежая информация, которая была использована для разработки проекта.</p> <p>С4. Невысокие системные требования.</p>	<p>Сл2. Ограниченный функционал конечного ПО.</p> <p>Сл3. Малый опыт создания подобных систем.</p>
<p><b>Возможности</b></p> <p>В1. Невысокий уровень конкуренции.</p> <p>В2. Появление потенциального спроса на новые разработки.</p>	<p><b>Направления развития</b></p> <p>В1С1С3. Возможности, предлагаемые проектом, и степень новизны используемой информации позволяют расширять рынок сбыта.</p>	<p><b>Сдерживающие факторы</b></p> <p>В1Сл3. Новизна технологии как таковой определяет невысокий уровень конкуренции.</p>
<p><b>Угрозы</b></p> <p>У1. Снижение стоимости разработок конкурентов.</p> <p>У2. Появление зарубежных аналогов и более ранний их выход на рынок.</p>	<p><b>Угрозы развития</b></p> <p>У1У2С2С4. Нетребовательность к оборудованию позволяет входить на рынок конкурентам. Противодействие угрозе заключается в более квалифицированном персонале и степени разработанности системы.</p>	<p><b>Уязвимости:</b></p> <p>У1Сл1. Введение систем совершенствования производственных процессов для снижения погрешности анализа.</p>

В результате выполнения SWOT-анализа можно увидеть что преимущества предлагаемого проекта преобладают над его недостатками, показано, что на преимущества разрабатываемой технологии преобладают над ее недостатками.

## 4.2 Планирование научно-исследовательских работ

### 4.2.1 Структура работ в рамках научного исследования

Планирование комплекса научно-исследовательских работ осуществляется в порядке:

- определение структуры работ в рамках научного исследования;
- определение количества исполнителей для каждой из работ;
- установление продолжительности работ;
- построение графика проведения научных исследований.

Для оптимизации работ удобно использовать классический метод линейного планирования и управления.

Результатом такого планирования является составление линейного графика выполнения всех работ. Порядок этапов работ и распределение исполнителей для данной научно-исследовательской работы, приведен в таблице 4.8.

Таблица 4.8 – Перечень этапов, работ и распределение исполнителей

Основные этапы	№ раб	Содержание работ	Должность исполнителя
Подготовка к анализу данных	1	Сбор информации, относящейся к теме исследования	Начальник отдела, Инженер
	2	Очистка и форматирование данных	Инженер
	3	Предварительный анализ данных	Инженер
Анализ данных	4	Выбор наиболее полезных признаков и создание новых более репрезентативных	Инженер
	5	Сравнение качества работы нескольких моделей	Начальник отдела, Инженер
	6	Проверка модели на тестовой выборке	Начальник отдела, Инженер

Обобщение и оценка результатов	7	Интерпретация результатов	Начальник отдела, Инженер
	8	Итоговое представление результатов работы	Начальник отдела

#### 4.2.2 Определение трудоемкости выполнения работ и разработка графика проведения

При проведении научных исследований основную часть стоимости разработки составляют трудовые затраты, поэтому определение трудоемкости проводимых работ является важным этапом составления сметы.

Для определения ожидаемого (среднего) значения трудоемкости использована следующая формула:

$$t_{\text{ож}i} = \frac{3t_{\text{мин}i} + 2t_{\text{макс}i}}{5}, \quad (4.1)$$

где  $t_{\text{ож}i}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы, человеко-дни;

$t_{\text{мин}i}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы, человеко-дни;

$t_{\text{макс}i}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы, человеко-дни.

Зная величину ожидаемой трудоемкости, можно определить продолжительность каждой  $i$ -ой работы в рабочих днях  $T_{pi}$ , при этом учитывается параллельность выполнения работ разными исполнителями. Данный расчёт позволяет определить величину заработной платы.

$$T_{pi} = \frac{t_{\text{ож}i}}{Ч_i}, \quad (4.2)$$

где  $T_{pi}$  – продолжительность одной работы, рабочие дни;

$t_{ожі}$  – ожидаемая трудоемкость выполнения одной работы, человеко-дни;

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

Для перевода длительности каждого этапа из рабочих в календарные дни, необходимо воспользоваться формулой (4.3):

$$T_{ki.инж} = T_{pi} \cdot k_{кал}, \quad (4.3)$$

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

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

$k_{кал}$  – календарный коэффициент.

Календарный коэффициент определяется по формуле:

$$k_{кал.инж} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}} = \frac{365}{365 - 104 - 14} = 1,48 \quad (4.4)$$

где  $T_{кал}$  – общее количество календарных дней в году;  $T_{вых}$  – общее количество выходных дней в году;  $T_{пр}$  – общее количество праздничных дней в году.

Расчеты временных показателей проведения научного исследования обобщены в таблице 4.9.

Таблица 4.9 – Временные показатели проведения научного исследования

Название работы	Трудоёмкость работ						Длительность работ в рабочих днях, $T_{pi}$	Длительность работ в календарных днях, $T_{ki}$
	$t_{min}$ , чел-дни		$t_{max}$ , чел-дни		$t_{ожі}$ , чел-дни			
	Исп.1	Исп.2	Исп.1	Исп.2	Исп.1	Исп.2		
1	2	3	4	5	6	7	8	9

Сбор информации, относящейся к теме исследования	4	10	6	14	4,8	11,6	8,2	12,1
Очистка и форматирование данных	-	2	-	5	-	3,2	3,2	4,7
Предварительный анализ данных	-	3	-	6	-	4,2	4,2	6,2
Выбор наиболее полезных признаков и создание новых более репрезентативных	-	2	-	4	-	2,8	2,8	4,1
Сравнение качества работы нескольких моделей	5	5	7	7	5,8	5,8	5,8	8,6
Проверка модели на тестовой выборке	2	2	4	4	2,8	2,8	2,8	4,1
Интерпретация результатов	4	4	6	6	4,8	4,8	4,8	7,1
Итоговое представление результатов работы	2	-	3	-	2,4	-	2,4	3,5
<b>Итого:</b>	<b>17</b>	<b>28</b>	<b>26</b>	<b>46</b>	<b>20,6</b>	<b>35,2</b>	<b>34,2</b>	<b>50,5</b>

Примечание: Исп. 1 – начальник отдела, Исп. 2 – инженер.

На основе таблицы составлен календарный план-график выполнения проекта с использованием диаграммы Ганта (таблица 4.10).

Таблица 4.10 – Диаграмма Ганта

№	Вид работ	Исп	$T_{ki}$ кал. дн.	Продолжительность работ							
				март			апрель			май	
				1	2	3	1	2	3	1	
1	Сбор информации, относящейся к теме исследования	Исп1 Исп2	12,1								
2	Очистка и форматирование данных	Исп2	4,7								

№	Вид работ	Исп	$T_{ki}$ кал. дн.	Продолжительность работ									
				март			апрель			май			
				1	2	3	1	2	3	1			
3	Предварительный анализ данных	Исп2	6,2			■							
4	Выбор наиболее полезных признаков и создание новых более репрезентативных	Исп2	4,1				■						
5	Сравнение качества работы нескольких моделей	Исп1 Исп2	8,6					▨	■				
6	Проверка модели на тестовой выборке	Исп1 Исп2	4,1							▨			
7	Интерпретация результатов	Исп1 Исп2	7,1								▨	■	
8	Итоговое представление результатов работы	Исп1	3,5										▨

Примечание:



– Исп. 1 (начальник отдела),



– Исп. 2 (инженер)

#### 4.2.3 Бюджет научно-технического исследования

При планировании бюджета научно-технического исследования учитывались все виды расходов, связанных с его выполнением. В этой работе использовать следующую группировку затрат по следующим статьям:

- материальные затраты научно-исследовательской работы (НИР);
- затраты на специальное оборудование для экспериментальных работ;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы НИР.

#### 4.2.3.1 Расчет материальных затрат научно-технического исследования

Материальные затраты — это затраты организации на приобретение сырья и материалов для создания готовой продукции.

Таблица 4.11 – Материальные затраты

Наименование материалов	Цена за ед., руб.	Кол-во, ед.	Сумма, руб.
Комплекс канцелярских принадлежностей	340	4	1 200
Картридж для лазерного принтера	3 490	1	3 490
Бумага для принтера	350	2	700
Итого:			8 990

Цены приняты на основании прайс-листа поставщика материалов: <https://papirus-tomsk.ru>, <https://mvideo.ru>.

#### 4.2.3.2 Расчет амортизации оборудования

Расчет сводится к определению амортизационных отчислений, так как оборудование было приобретено до начала выполнения данной работы и эксплуатировалось ранее, поэтому при расчете затрат на оборудовании учитываем только рабочие дни по данной теме.

Таблица 4.12 – Затраты на оборудование

№	Наименование оборудования	Кол-во, шт.	Срок полезного использования, лет	Цены единицы оборудования, тыс. руб.	Общая стоимость оборудования, тыс. руб.
1	ПЭВМ	2	3	87	174
<b>Итого</b>				174 тыс. руб.	

Расчет амортизации проводится следующим образом:

Норма амортизации: рассчитывается по формуле:

$$H_A = \frac{1}{n}, \quad (4.5)$$

где  $n$  – срок полезного использования в количестве лет.

Амортизация оборудования рассчитывается по формуле:

$$A = \frac{H_A I}{12} \cdot m, \quad (4.6)$$

где  $I$  – итоговая сумма, тыс. руб.;  $m$  – время использования, мес.

Расчет нормы амортизации оборудования из таблицы 4.12:

$$H_A = \frac{1}{n} = \frac{1}{3} = 0,33.$$

Общую сумму амортизационных отчислений находим следующим образом:

$$A = \frac{H_A I}{12} \cdot m = \frac{0,33 \cdot 174000}{12} \cdot 1,68 = 8\,038,8 \text{ руб.}$$

#### 4.2.3.3 Заработная плата исполнителей темы

В данном разделе рассчитывается заработная плата инженера и руководителя, помимо этого необходимо рассчитать расходы по заработной плате, определяемые трудоемкостью проекта и действующей системой оклада.

Заработная плата  $Z_{осн}$  одного работника рассчитывается по следующей формуле:

$$Z_{осн} = Z_{он} \cdot T_p, \quad (4.7)$$

где  $Z_{\text{дн}}$  – среднедневная заработная плата, руб.;  $T_p$  – продолжительность работ, выполняемых работником, раб.дн. (таблица 4.9).

Среднедневная заработная плата руководителя рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_m \cdot M}{F_{\text{д}}} = \frac{138866,2 \cdot 10,2}{247} = 5734,6 \text{ руб.}, \quad (4.8)$$

где  $Z_m$  – месячный должностной оклад работника, руб.;  $F_{\text{д}}$  – действительный годовой фонд рабочего времени, раб. дней;  $M$  – количество месяцев работы без отпуска в течение года:

– при отпуске в 58 раб. дней –  $M = 10,2$  месяца.

Среднедневная заработная плата инженера рассчитывается по формуле:  $Z_{\text{дн}} =$

$$\frac{Z_m \cdot M}{F_{\text{д}}} = \frac{107\,641,6 \cdot 10,6}{247} = 4619,4 \text{ руб.} \quad (4.9)$$

– при отпуске в 47 раб. дней –  $M = 10,6$  месяца.

Должностной оклад работника за месяц:

– для руководителя:

$$Z_m = Z_{\text{мс}} \cdot (1 + k_{\text{нр}} + k_{\text{д}})k_p = 37130 \cdot (1 + 0,4 + 0,3) \cdot 2,2 = 138\,866,2 \text{ руб.} \quad (4.10)$$

– для инженера:

$$Z_m = Z_{\text{мс}} \cdot (1 + k_{\text{нр}} + k_{\text{д}})k_p = 30580 \cdot (1 + 0,4 + 0,2) \cdot 2,2 = 107\,641,6 \text{ руб.}, \quad (4.11)$$

где  $Z_{\text{мс}}$  – заработная плата, согласно тарифной ставке, руб.;  $k_{\text{нр}}$  – премиальный коэффициент, равен 0,4;  $k_{\text{д}}$  – коэффициент доплат и надбавок;

$k_p$  – районный коэффициент и северная надбавка, равен 2,2 (для г. Стрежевого).

Таблица 4.13 – Баланс рабочего времени исполнителей

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней	104/14	104/14
- выходные дни		
- праздничные дни		
Потери рабочего времени	58/5	47/5
- отпуск		
- невыходы по болезни		
Действительный годовой фонд рабочего времени	184	195

Таблица 4.14 – Расчет основной заработной платы исполнителей

Исполнители НИ	$Z_{мс}, руб$	$k_{np}$	$k_{\delta}$	$k_p$	$Z_m, руб$	$Z_{осн}, руб$	$T_p, раб.дн.$	$Z_{осн}, руб$
Руководитель	37130	0,4	0,3	2,2	138 866	5 734,6	20,6	118 132,8
Инженер	30580	0,4	0,2	2,2	107 642	4 619,4	35,2	162 602,9
Итого:								280 735,7

#### 4.2.3.4 Дополнительная заработная плата исполнителей темы

Дополнительная заработная плата определяется по формуле:

– для руководителя:

$$Z_{доп} = k_{доп} \cdot Z_{осн} = 0,15 \cdot 118132,8 = 17\,719,9 \text{ руб.} \quad (4.12)$$

– для инженера:

$$Z_{доп} = k_{доп} \cdot Z_{осн} = 0,15 \cdot 162602,9 = 24\,390,4 \text{ руб.} \quad (4.13)$$

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

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

$$З_{внеб} = k_{внеб} \cdot (З_{осн} + З_{доп}),$$

Отчисления во внебюджетные фонды определяется по формуле:

– для руководителя:

$$З_{внеб} = k_{внеб} \times (З_{осн} + З_{доп}) = 0,3 \times (118132,8 + 17719,9) = 40\,755,8 \text{ руб.} \quad (4.14)$$

– для инженера:

$$З_{внеб} = k_{внеб} \times (З_{осн} + З_{доп}) = 0,3 \times (162602,9 + 24390,4) = 56\,098 \text{ руб.}, \quad (4.15)$$

где  $k_{внеб}$  – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд ОМС и социальное страхование). Общая ставка взносов составляет в 2020 году – 30% (ст. 425, 426 НК РФ).

#### 4.2.3.6 Накладные расходы

Накладные расходы представляют собой затраты на управление, организацию и обслуживание производства и применяются коэффициентом ко всем статьям затрат. В данном случае, коэффициент примем равным 0,2 и составим калькуляцию себестоимости разработки модели анализа данных.

Таблица 4.16 – Группировка затрат по статьям

№	Наименование статьи	Сумма, руб.		Примечание
		Проект	Вар.2	
1	Материальные затраты	8 990	9 990	Пункт 4.2.3.1

2	Затраты на оборудование	8 038	12 057	Пункт 4.2.3.2
3	Заработная плата исполнителей проекта	280 736	443 339	Пункт 4.2.3.3
4	Дополнительная заработная плата	42 110	66 501	Пункт 4.2.3.4
5	Отчисления во внебюджетные фонды	96 584	152 952	Пункт 4.2.3.5
6	Накладные расходы	87 292	136 968	Пункт 4.2.3.6
Итого		523 750	821 807	Сумма ст. 1- 6

#### **4.3 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования**

Для определения эффективности исследования рассчитан интегральный показатель эффективности научного исследования путем определения интегральных показателей финансовой эффективности и ресурсоэффективности.

**Интегральный показатель финансовой эффективности** научного исследования получен в процессе оценки затрат двух вариантов исполнения научного исследования. Для этого наибольший интегральный показатель реализации технической задачи принят за базу расчета (как знаменатель), с которым соотносятся финансовые значения по всем вариантам исполнения.

В качестве аналога данному проекту рассмотрена замена насосно-компрессорных труб на основе графиков планово-предупредительных ремонтов;

Интегральный финансовый показатель разработки рассчитывается как:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}, \quad (4.17)$$

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

$\Phi_{pi}$  – стоимость  $i$ -го варианта исполнения;

$\Phi_{\text{max}}$  – максимальная стоимость исполнения.

$\Phi_{\text{проект}} = 523\,750$  руб.,  $\Phi_{\text{вар.2}} = 821\,807$  руб.

$$I_{\text{финр}}^{\text{тек.пр.}} = \frac{\Phi_{\text{тек.пр.}}}{\Phi_{\text{max}}} = \frac{523\,750}{821\,807} = 0,64;$$

$$I_{\text{финр}}^{\text{исп.2}} = \frac{\Phi_{\text{вар.2}}}{\Phi_{\text{max}}} = \frac{821\,807}{821\,807} = 1.$$

В результате расчета консолидированных финансовых показателей по двум вариантам разработки вариант 1 (текущий проект) признан более приемлемым с точки зрения финансовой эффективности.

**Интегральный показатель ресурсоэффективности** вариантов выполнения НИР ( $I_{pi}$ ) определен путем сравнительной оценки их характеристик, распределенных с учетом весового коэффициента каждого параметра (таблица 4.17).

Таблица 4.17 – Сравнительная оценка характеристик вариантов НИР

Критерии \ Объект исследования	Весовой коэффициент параметра	Текущий проект	Вар.2
Актуальность	0,3	5	4
Удобство в эксплуатации	0,2	5	2
Простота разработки	0,2	2	5
Эффективность работы	0,15	5	3
Надежность продукта	0,15	4	2

ИТОГО	1	4,65	3,8
-------	---	------	-----

Расчет интегрального показателя для разрабатываемого проекта:

$$I_{p1} = 0,3 \cdot 5 + 0,2 \cdot 5 + 0,2 \cdot 2 + 0,15 \cdot 5 + 0,15 \cdot 4 = 4,25;$$

$$I_{p2} = 0,3 \cdot 4 + 0,2 \cdot 2 + 0,2 \cdot 5 + 0,15 \cdot 3 + 0,15 \cdot 2 = 3,35;$$

Интегральный показатель эффективности вариантов исполнения разработки вычисляется на основании показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.i} = \frac{I_{p-исп.i}}{I_{финр.i}} \quad (20)$$

$$I_{проект} = \frac{4,25}{0,64} = 6,64, \quad I_{вар.2} = \frac{3,35}{1} = 3,35.$$

Далее интегральные показатели эффективности каждого варианта НИР сравнивались с интегральными показателями эффективности других вариантов с целью определения сравнительной эффективности проекта (таблица 4.18).

Таблица 4.18 – Сравнительная эффективность разработки

№ п/п	Показатели	Текущий проект	Вар.2
1	Интегральный финансовый показатель разработки	0,64	1
2	Интегральный показатель ресурсоэффективности разработки	4,25	3,35
3	Интегральный показатель эффективности	6,64	3,35
4	Сравнительная эффективность вариантов исполнения	1	0,67

Сравнение среднего интегрального показателя сопоставляемых вариантов позволило сделать вывод о том, что наиболее финансово- и ресурсоэффективным является вариант 1 (текущий проект).

## **Выводы по разделу**

В результате выполнения целей раздела можно сделать следующие выводы:

1. Результатом анализа конкурентных технических решений является выбор одного из вариантов реализации НИР как наиболее подходящего и оптимального по сравнению с другими.

2. В ходе планирования для руководителя и инженера был разработан график реализации этапа работ, который позволяет оценивать и планировать рабочее время исполнителей. Определено следующее: общее количество дней для выполнения работ составляет 50,5 дней; общее количество дней, в течение которых работал инженер, составляет 35,2 дней; общее количество дней, в течение которых работал руководитель, составляет 20,6 дней;

3. Для оценки затрат на реализацию проекта разработан проектный бюджет, который составляет 523 750 руб;

4. Результат оценки эффективности ИР показывает следующие выводы:

1) значение интегрального финансового показателя ИР составляет 0,64, что является показателем того, что ИР является финансово выгодной в сравнении с аналогом;

2) значение интегрального показателя ресурсоэффективности ИР составляет 4,25, что значительно выше показателя варианта №2 в 3,35;

3) значение интегрального показателя эффективности ИР составляет 6,64 и является наиболее высоким, что означает, что техническое решение, рассматриваемое в ИР, является наиболее эффективным вариантом исполнения;

## 5. Социальная ответственность

### Введение

В данной главе освещен комплекс мер организационного, правового, технического и режимного характера, которые минимизируют негативные последствия разработки программного комплекса, а также рассматриваются вопросы техники безопасности, охраны окружающей среды и пожарной профилактики, даются рекомендации по созданию оптимальных условий труда.

Объектом исследования выступает рабочее место программиста, разрабатывающего данную методологию, которая позволит корректно обрабатывать исходные данные с нефтегазового месторождения.

Рабочее место – это часть пространства в офисе, в котором сотрудник осуществляет трудовую деятельность. Помещение, в котором находится рабочее место, имеет следующие характеристики:

- длина помещения (А): 6 м;
- ширина помещения (В): 5 м;
- высота помещения (Н): 3 м;
- число окон: 2 (размер 1,5x2 м);
- число рабочих мест: 2.

В помещении имеются рабочие места, оснащенные компьютерами и принтером.

Главными элементами рабочего места являются стол и кресло. Рабочий стол имеет пространство для ног высотой 700 мм, шириной 600 мм, глубина на уровне колен – 450 мм и на уровне вытянутых ног – 700 мм.

Помещение характеризуется как объект с малым выделением пыли. Естественная вентиляция осуществляется при помощи окон и вентиляционных каналов. Отопление осуществляется посредством системы центрального водяного отопления.

## **5.1. Правовые и организационные вопросы обеспечения безопасности**

### **5.1.1. Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства.**

Правовое регулирование трудовых отношений между работодателем, работником и государством регулируется Трудовым кодексом Российской Федерации от 30.12.2001 N 197-ФЗ. В ТК РФ, в соответствии с Конституцией РФ, признаются свобода труда, выбор и согласие на него, а также выбор профессии и деятельности. Запрещаются принудительный труд, дискриминация по какому-либо признаку. Гарантируются справедливые и достойные условия труда.

ТК РФ регламентирует порядок разрешения индивидуальных и коллективных трудовых споров, особенности труда женщин, детей и людей пенсионного возраста, права и обязанности работодателей и работников, нормы рабочего времени, порядок оплаты труда и виды компенсаций во вредных условиях труда, а также особенности социального страхования.

В соответствии со ст. 111 ТК РФ, рабочая неделя (в т.ч., шестидневная) не должна превышать 40 часов в неделю. Воскресенье является выходным днем.

В соответствии со ст. 212 ТК РФ, работодатель обязан обеспечить безопасные условия труда, а также обязательное социальное страхование работников от несчастных случаев на производстве и профессиональных заболеваний.

В соответствии со ст. 142 ТК РФ, в случае задержки выплаты заработной платы на срок более 15 дней работник имеет право, известив работодателя в письменной форме, приостановить работу на весь период до выплаты задержанной суммы, кроме ряда перечисленных случаев.

Данные о работнике, предоставляемые работодателю, обрабатываются только с согласия самого работника и охраняются

Федеральным Законом от 27.07.2006 N 152-ФЗ (ред. от 25.07.2011) «О  
Персональных Данных».

### **5.1.2. Организационные мероприятия при компоновке рабочей зоны**

Согласно ГОСТ 12.2.032-78 ССБТ, выявлены следующие параметры рабочей зоны:

- Согласно наименованию работы (работа за ЭВМ) при отсутствии регулирующих механизмов высоты рабочей поверхности, высота рабочей поверхности, при организации рабочего места, составляет (для мужчин) 700 мм. Высота сиденья 500 мм.
- Рабочая поверхность в соответствии с видом работ может содержать дополнительное углубление для периферийных устройств (клавиатура).
- Рабочее место при выполнении работ сидя организуют при легкой работе, не требующей свободного передвижения работающего.
- Конструкция рабочего места и взаимное расположение всех его элементов должны соответствовать антропометрическим, физиологическим и психологическим требованиям, а также характеру работы.

В соответствии с СП 2.2.3670-20 "Санитарно-эпидемиологические требования к условиям труда" были выявлены следующие правила организации работы с ПЭВМ:

- Площадь на одно постоянное рабочее место пользователей персональных компьютеров на базе электронно-лучевой трубки, должна составлять не менее 6 м, в помещениях культурно-развлекательных учреждений, на базе плоских дискретных экранов (жидкокристаллические, плазменные) - не менее 4,5 м.

**5.2.** Персональные компьютеры следует размещать таким образом, чтобы показатели освещенности не превышали установленных гигиенических нормативов утвержденных в соответствии с пунктом 2 статьи 38 Федерального закона от 30.03.1999 N 52-ФЗ "О санитарно-эпидемиологическом благополучии населения".

### 5.3. Производственная безопасность

В данном подразделе анализируются вредные и опасные факторы, которые могут возникать при проведении исследований в лаборатории, при разработке или эксплуатации проектируемого решения.

В таблице 5.1 представлены возможные вредные и опасные факторы. Работа по разработке программного обеспечения делится на три основных этапа: проектирование, разработка и эксплуатация.

Таблица 5.1. Возможные вредные и опасные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ	Нормативные документы
	Разработка	
Отклонение показателей микроклимата	+	ГОСТ 12.1.005-88 ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны  СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания.
Превышение уровня шума	+	ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности
Недостаточная освещенность рабочей зоны	+	СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95

<p>Опасность поражения электрическим током</p>	<p>+</p>	<p>ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ). Электробезопасность. Защитное заземление. Зануление.</p> <p>ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов</p>
--	----------	--

### **5.2.1. Анализ вредных и опасных факторов, которые может создать объект исследования.**

Теоретически, объект исследования при определенных условиях, не предусмотренных разработчиками, способен привести к перегрузкам в ПЭВМ и вызвать определенные последствия. Поскольку разработка не является осязаемым объектом и неотделима от ПЭВМ, вредные и опасные факторы, которые могут быть прямо или косвенно отнесены к разработке, относятся и к рабочему месту.

### **5.2.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте.**

#### **5.2.2.1. Отклонение показателей микроклимата.**

Микроклимат определяется такими показателями, как влажность и температура воздуха. Поддержка их в установленных общими санитарно-гигиеническими требованиями к воздуху рабочей зоны пределах, согласно ГОСТ 12.1.005-88 ССБТ, осуществляется при помощи общей вентиляции (открывающиеся окна) и системы отопления помещения. Также периодически проводится контроль влажности воздуха.

При энергозатратах до 120 ккал/час (до 139 Вт) выполняемая работа

относится к категории 1а – производится сидя, сопровождается незначительным физическим напряжением. Оптимальные нормы микроклимата для данной категории приведены в таблице 5.2.

Таблица 5.2. 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

Если оптимальные нормы микроклимата не могут быть обеспечены, устанавливаются допустимые значения показателей микроклимата (представлены в таблице 5.2) в соответствии с требованиями СанПиН 1.2.3685-21.

Таблица 5.2. Допустимые значения показателей микроклимата

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

В помещении, где велась разработка, использовались доступные способы поддержания показателей микроклимата в допустимых пределах. В холодные периоды года использовалась система водяного центрального отопления. В теплые периоды и при повышении температуры воздуха в помещении комфортные условия достигались при помощи системы вентиляции.

При температуре воздуха 19-22 °С и относительной влажности 55-60% показатели микроклимата удовлетворяют требованиям ГОСТ

12.1.005-88 ССБТ.

#### **5.2.2.2. Превышение уровня шума**

Результатом воздействия шума на человека может быть повышенная утомляемость, снижение внимания и ухудшение работы органов слуха, что, в свою очередь, приводит к ухудшению работоспособности (увеличение затраченного времени). Согласно ГОСТ 12.1.003-2014, нормативный уровень шума на рабочем месте составляет 80 дБА.

На используемом рабочем месте основным источником шума является персональный компьютер. Согласно документации ПК, уровень шума вентилятора центрального процессора и вентилятора блока питания не превышает 40 дБА (используемый ПК классифицируется как тонкий клиент). В комплектацию ПК не входят высокопроизводительная видеокарта или дополнительное охлаждение. Данный источник шума не превышает нормативных показателей.

Источники внешнего шума могут достигать значений 65-70 дБА (открытое окно), средний показатель 52 дБА. Снижение уровня шума до 47 дБА было достигнуто путем использования занавесок из плотной ткани.

Рабочее место, на котором проводилась разработка, не превышает допустимых значений по уровню шума и соответствует установленным требованиям ГОСТ 12.1.003-2014. Замеры производились при помощи мобильного приложения «Шумомер и детектор шума».

#### **5.2.2.3. Недостаточная освещенность рабочей зоны**

Одним из основных факторов, влияющих на пригодность рабочего места к эксплуатации, является освещение. Оно может быть как естественным, так и искусственным. Естественное освещение обеспечивается через оконные проемы, световой поток при этом должен падать на рабочее место с левой стороны от оператора. Коэффициент естественного освещения (КЕО) оконных проемов должен быть не менее

1,5% или 1,2% в местах со снежным покровом.

Освещенность рабочего места определяется с таким расчетом, чтобы исключить прямое попадание лучей в глаза и обеспечить отчетливое различение объектов рабочей поверхности оператору без необходимости напрягать зрение. При недостаточной освещенности используются искусственные источники света.

Работа за ПК относится к зрительным работам высокой точности. В таблице 5.3 приведены требования к показателям освещенности при работе за ПК.

Таблица 5.3. Нормы освещения для работы за ПК

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

#### 5.2.2.4. Опасность поражения электрическим током

ГОСТ 12.1.038-82 определяет предельно допустимые значения напряжений прикосновения и токов, протекающих через тело человека при продолжительности воздействия более одной секунды. Разработка проекта не предполагает взаимодействие с открытыми проводниками электрического тока.

Платы микросхем используемой ЭВМ закрыты стандартным корпусом. Провода питания компьютера и периферических устройств периодически проверяются на наличие неисправностей изоляции. Питание осуществляется через общую электрическую сеть с заземлением согласно требованиям ГОСТ 12.1.030-81. Используется промежуточный фильтр с автоматическим предохранителем, защищающий устройства от резких

перепадов напряжения.

#### **5.4. Экологическая безопасность**

Целью данного подраздела является выявление потенциальных опасностей объекта и процесса исследования на окружающую среду, а также разработка мер, обеспечивающая безопасность исследовательской деятельности для окружающей среды.

##### **5.3.1. Анализ влияния объекта исследования на окружающую среду**

Объект исследования (методология подготовки исходных данных) не оказывает влияния на окружающую среду, поскольку используется только совместно с ПЭВМ. Сами ПЭВМ могут являться источниками различных загрязнений окружающей среды.

##### **5.3.2. Анализ влияния процесса исследования на окружающую среду**

Процесс исследования включает в себя работу на ПЭВМ в офисном помещении, в том числе в условиях искусственного освещения, обеспечиваемого люминесцентными лампами.

Отработанная офисная техника относится к опасным отходам. При производстве компьютеров и других агрегатов применяются вещества, опасные для жизнедеятельности, например, свинец, мышьяк и др. Обычное выбрасывание техники, особенно регулярное, может нанести непоправимый вред экологии и здоровью. Согласно Административному Кодексу РФ, ст. 8.2, запрещается выбрасывать технику наряду с обычным мусором, причем запрет распространяется не только на физических лиц, но и на организации.

##### **5.3.3. Обоснование мероприятий по защите окружающей среды**

Согласно ГОСТ Р 56397-2015, в результате технической экспертизы может быть принято следующее решение: оборудование не ремонтпригодно, признается неработоспособным и рекомендуется к

списанию (замене); в случае деградационного отказа оборудования и нецелесообразности его ремонта и модернизации даются рекомендации о необходимости его списания и утилизации. Самостоятельная утилизация оргтехники запрещена, утилизация производится только в промышленных условиях. Утилизировать компьютерную технику имеют права специализированные предприятия при наличии соответствующей лицензии.

На рабочем месте программиста используются 16 люминесцентных ламп ЛБ40, Согласно ГОСТ 12.3.031-83 «Работы со ртутью. Требования безопасности» п.2.1. все ртутьсодержащие отходы и вышедшие из строя приборы, содержащих ртуть, подлежат сбору и возврату для последующей регенерации ртути в специализированных организациях. Согласно п.2.2 к работе по замене и сбору отработанных ртутьсодержащих ламп допускаются только электромонтеры. Главным условием при замене и сборе отработанных ртутьсодержащих ламп является сохранение герметичности.

#### **5.3.4. Безопасность в чрезвычайных ситуациях**

#### **5.3.5. Анализ вероятных ЧС, которые может инициировать объект исследований**

В ходе проведения анализа не было выявлено ЧС, которые может инициировать объект исследования напрямую.

#### **5.3.6. Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований**

К наиболее вероятным ЧС на рабочем месте можно отнести следующие: пожар (взрыв) в здании, авария на коммунальных системах жизнеобеспечения, землетрясение.

Наиболее вероятным ЧС является пожар. Источниками возгорания может стать электропроводка, внутренние работающие устройства ПК.

Поражающими факторами пожаров в помещении являются

токсическое воздействие горючих материалов (в т.ч. отравление угарным газом), экстремальный нагрев среды, а также обломки и осколки при нарушении целостности конструкций здания.

### **5.3.7. Обоснование мероприятий по предотвращению ЧС и разработка порядка действий в случае возникновения ЧС**

Согласно ГОСТ Р 22.3.03-94, обеспечение безопасности людей в ЧС, обусловленных природными стихийными бедствиями, техногенными авариями и катастрофами, а также применением современного оружия (военные ЧС) является общегосударственной задачей, обязательной для решения всеми территориальными, ведомственными и функциональными органами управления и регулирования, службами и формированиями, а также подсистемами, входящими в Российскую систему предупреждения и действий в чрезвычайных ситуациях (РСЧС).

Мероприятия по защите людей от источников ЧС должны планироваться в объемах, гарантирующих не превышение нормативного воздействия на них возможных поражающих факторов для расчетной ЧС.

Для защиты жизни и здоровья населения в ЧС следует применять следующие основные мероприятия гражданской обороны, являющиеся составной частью мероприятий РСЧС:

1) укрытие людей в приспособленных под нужды защиты населения помещениях производственных, общественных и жилых зданий, а также в специальных защитных сооружениях;

2) эвакуацию населения из зон ЧС;

3) использование средств индивидуальной защиты органов дыхания и кожных покровов;

4) проведение мероприятий медицинской защиты;

5) проведение аварийно-спасательных и других неотложных работ в зонах ЧС.

Мерами по предупреждению ЧС являются:

1) Соблюдение техники безопасности при работе с ПЭВМ. Использование только исправного оборудования.

2) Своевременное проведение ТО и ППР электроустановок согласно утвержденного графика и технических средств противопожарной защиты и пожаротушения.

3) Установка противопожарной сигнализации.

4) Своевременное проведение инструктажа рабочего персонала.

В случае угрозы возникновения ЧС (пожара) необходимо вызвать противопожарную службу, отключить электроэнергию и, следуя плану эвакуации, эвакуировать находящихся в помещении людей и покинуть помещение. В случае, если очаг возгорания является небольшим, и нет угрозы поражения электрическим током, можно использовать углекислотные огнетушители ОУ-5 высокого давления с зарядом жидкой двуокиси углерода, согласно ГОСТ 8050-85.

### **Вывод по главе**

В данной главе были проанализированы опасные и вредные факторы труда разработчика, а также предложены меры защиты от них, оценены условия труда рабочей зоны; рассмотрены требования по технике безопасности, электробезопасности, пожарной безопасности, экологической безопасности.

В ходе анализа выявлено, что рабочее помещение относится к 1 категории по электробезопасности, категории Д по взрывопожарной и пожарной опасности, персонал относится к 1 группе по электробезопасности и к категории работ 1а по тяжести труда.

В результате анализа было установлено, что офисное помещение, где проводится разработка проектного решения, удовлетворяет всем требованиям, предъявляемым к нему нормативными документами в

области охраны труда и окружающей природной среды.

## ЗАКЛЮЧЕНИЕ

В результате выполнения данной работы достигнута поставленная цель, в частности: исследованы методы анализа данных для прогнозной аналитики в нефтегазовой компании на основе методов машинного обучения для прогнозирования отказов насосно-компрессорных труб.

В процессе достижения поставленной цели решены следующие задачи:

- Выполнен анализ актуальности и проблематики нефтегазовой сферы деятельности.
- Проведен анализ возможностей и преимуществ использования современных методов интеллектуального анализа данных.
- Осуществлен анализ особенностей прогнозирования данных.
- Проведен обзор методов и моделей функционирования искусственных нейронных сетей.
- Выполнен анализ существующих программных продуктов и систем проведения интеллектуального анализа данных.
- Реализован анализ возможностей языков программирования для задач анализа данных.
- Осуществлен анализ ключевых преимуществ программных библиотек поддержки процессов анализа данных.
- Построены UML-диаграммы для отображения статической структуры системы, поведенческих и физических аспектов системы.
- Разработана обобщённая функциональная структура системы прогнозирования.
- Выполнена формализация концепции проводимых исследований.
- Описан набор данных для анализа.
- Проведены обработка и анализ данных с использованием методов машинного обучения.

## СПИСОК ЛИТЕРАТУРЫ

1. Худякова О. Ю., Исмаилова А.М. Современные тренды развития мирового рынка нефти / Научные междисциплинарные исследования. – №4. – 2021. – С. 143-150.
2. Хасанов А.Р. Влияние предиктивной аналитики на деятельность компаний / Стратегические решения и риск-менеджмент. – 2018. – №3. – С. 108-113.
3. Особенности нефтегазовой отрасли // АО «ЭКСПОЦЕНТР». [2022]. URL: <https://www.neftegaz-expo.ru/ru/articles/osobennosti-neftegazovoj-otrasli/> (дата обращения: 05.05.2022).
4. Современные тенденции развития нефтегазового комплекса: монография / О.В. Пожарницкая, И.В. Шарф, М.Р. Цибульникова. – Томск: Изд-во Томского политехнического университета, 2015. – 166 с.
5. Калашникова И.В., Воприкова А.А. Состояние и проблемы развития нефтегазового комплекса России / Электронное научное издание «Ученые заметки ТОГУ». – 2015. – Т. 6. – №2. – С.152 – 156
6. Каралкин М.В. Анализ состояния и возможности развития нефтегазового комплекса России / *π-Economy*. – 2011. – №2 (119). – С. 50-55.
7. Меренков М.Г., Никаноров Р.О., Рутелевская Ю.Д. Нефтегазовая отрасль России: проблемы и пути их решения / Актуальные проблемы авиации и космонавтики. – 2017. – Т. 3. – №13. – С. 66-68.
8. Чукомина Д.А. Основные проблемы развития нефтегазовой отрасли в Российской Федерации / Экономика нового мира. – 2020. – Т. 5. – №1-2 (17). – С. 47-54.
9. Сибгатуллина А.Я. Проблемы внедрения систем энергетического менеджмента на предприятиях нефтегазовой отрасли / Вестник магистратуры. – 2018. – №3(78). – С. 97-99.
10. Карамсейли А.А. Новые технические решения в нефтегазовой отрасли: проблемы и решения использования контейнерных АЗС / Инновации и инвестиции. – 2019. – №1. – С. 237-240.

11. Насосно-компрессионные трубы и их назначение и маркировка. Эксплуатация // Neftegaz.RU. [2000-2022]. URL: <https://neftegaz.ru/tech-library/burovye-ustanovki-i-ikh-uzly/141509-nasosno-kompressornye-truby-ikh-naznachenie-i-markirovka-ekspluatatsiya/> (дата обращения: 05.05.2022).

12. Ветров Д.П. Машинное обучение - состояние и перспективы / Электронные библиотеки: перспективные методы и технологии, электронные коллекции. XV Всероссийская конференция RCDL'2018. – Т. 1. – Издательство ЯрГУ Ярославль, 2018. – С. 21–28.

13. Чубукова И.А. Data Mining. – НОУ «Интуит», 2016. – 471 с.

14. Мулюкова К.В. Сравнительный анализ современных инструментов Data Mining // Молодой ученый. – 2019. – №1. – С. 19-21

15. Биктимиров М.Р. Тенденции развития технологий обработки больших данных и инструментария хранения разноформатных данных и аналитики // Russian Digital Libraries Journal. – 2016. – № 5. – Т.19. – С.390-406.

16. Голенков В.В. Традиционные и интеллектуальные информационные технологии. – Минск: БГУИР, 2016. – 64 с.

17. Миркин Б.Г. Введение в анализ данных. Учебник и практикум для бакалавриата и магистратуры. – М.: Юрайт, 2015. – 174 с.

18. Пальмов С.В. Интеллектуальный анализ данных. – Самара: Изд-во ПГУТИ, 2017. – 128 с.

19. Рафалович В. Data mining, или Интеллектуальный анализ данных для занятых. Практический курс. – М.: СмартБук, 2014. – 110 с.

20. Храмов А.Г. Методы и алгоритмы интеллектуального анализа данных. – Самара: Самарский национальный исследовательский университет им. академика С.П. Королева, 2019. – 176 с.

21. Основы теории прогнозирования // svetunkov.ru. URL: <https://forecasting.svetunkov.ru/etextbook/about/> (дата обращения: 05.05.2022).

22. Тихонов Э.Е. Прогнозирование в условиях рынка / Э.Е. Тихонов. – Невинномысск: Образование, 2016. – 221 с.

23. Снитюк В.Е. Прогнозирование. Модели, Методы, Алгоритмы / В.Е. Снитюк. – Киев: Маклаут, 2018. – 367 с.
24. Любимова Т.В. Решение задачи прогнозирования с помощью нейронных сетей / Т.В. Любимова, А.В. Горелова // Инновационная наука. – 2015. – №5. – С. 39-43.
25. Галушкин А.И. Нейронные сети: основы теории. – М.: Горячая линия - Телеком, 2018. – 496 с.
26. Панфилов П.Н. Введение в нейронные сети // Современный трейдинг. – 2011. – № 2. – С. 12–17.
27. Дунин-Барковский В.Л. Нейроинформатика. – М.: ИНТУИТ, 2016. – 330 с.
28. Ключко В.И. Нейросетевые топологии с подкреплением. – Краснодар: КубГТУ, 2017. – 154 с.
29. Николенко С. Глубокое обучение. Погружение в мир нейронных сетей. – СПб.: Питер, 2018. – 480 с.
30. Рашид Т. Создаем нейронную сеть. – СПб.: Альфа-книга, 2017. – 274 с.
31. Тадеусевич Р. Основы нейрокибернетики. – М.: Горячая линия - Телеком, 2015. – 372 с.
32. Холодов Г.М. Алгоритмическое и объектно-ориентированное программирование. – М.: МГТУ МАМИ, 2018. – 459 с.
33. Хорев П.Б. Объектно-ориентированное программирование. – Москва: Академия, 2015. – 446 с.
34. Складнев О.М. Разработка программных продуктов на языках высокого уровня. – СПб.: Глоунас, 2014. – 541 с.
35. Камаев В.А. Технологии программирования. – М.: Высш. шк., 2016. – 314 с.
36. Нейгел К., Ивсен Б., С# 5.0 и платформа .NET 4.5 для профессионалов. М.: Академия, 2014. – 1440 с.

37. Рихтер Д. Программирование на платформе Microsoft .NET Framework 4.0 на языке C#. – СПб.: Питер, 2013. – 720 с.
38. Федоров А. Г. Создание Windows-приложений на C#. – М.: ТОО «Компьютер Пресс», 2014. – 297 с.
39. Закас Н. JavaScript для профессиональных веб-разработчиков. – СПб.: Питер, 2015. – 960 с.
40. Хэррон Д. Node.js Разработка серверных веб-приложений на JavaScript. – СПб.: ДМК Пресс, 2015. – 518 с.
41. Васильев А. Программирование на JavaScript в примерах и задачах. М.: Эксмо, 2017. - 231 с.
42. Буйначев С.К. Основы программирования на языке Python г. – М.: Издательство Уральского университета, 2017. – 92 с.
43. Прохоренок Н.А. PyQt. Создание оконных приложений на Python 3. – М.: Мир, 2012. – 317 с.
44. Сузи Р.А. Язык программирования Python. – М.: Интуит, 2016. – 351 с.
45. Васильев А.Н. Python на примерах. Практический курс по программированию. – М.: Наука и техника, 2016. – 432 с.
46. Саммерфилд М. Программирование на Python 3. – СПб.: Символ-Плюс, 2017. – 718 с.
47. Лутц М. Изучаем Python. – СПб.: Символ-Плюс Медиа, 2019. – 1280 с.
48. Чан У. Python: создание приложений. – СПб.: Вильямс, 2015. – 612 с.
49. Keras – последовательная модель Sequential // proproprogs.ru. [2022]. URL: <https://proproprogs.ru/tensorflow/keras-posledovatel'naya-model-sequential> (дата обращения: 05.05.2022).

## ПРИЛОЖЕНИЕ А. ЛИСТИНГ КОДА

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
# %matplotlib inline

dataset = pd.read_csv('DatasetFinal.csv', delimiter=';',
decimal=',')

dataset.shape

dataset.head()

# fixes several incorrect values and turns column into
floats
dataset['Типоразмер      НКТ'].replace({'48*60':      54},
inplace=True)
dataset['Типоразмер      НКТ'].replace({'48*73':      59},
inplace=True)
dataset['Типоразмер      НКТ']      =      dataset['Типоразмер
НКТ'].astype(float)

dataset.isna().sum()

# заполнение пустых мест
dataset['Н      спуска'].fillna(dataset['Н      спуска'].mean(),
inplace=True)
dataset['Вид ремонта      НКТ'].fillna(dataset['Вид ремонта
НКТ'].mode()[0], inplace=True)
dataset['Завод-изготовитель      (трубная
база)'].fillna(dataset['Завод-изготовитель      (трубная
база)'].mode()[0], inplace=True)
dataset['Исполнение      НКТ'].fillna(dataset['Исполнение
НКТ'].mode()[0], inplace=True)
dataset['Типоразмер      НКТ'].fillna(dataset['Типоразмер
НКТ'].mode()[0], inplace=True)
dataset['Группа прочности      НКТ'].fillna(dataset['Группа
прочности НКТ'].mode()[0], inplace=True)
dataset['Кол-во      СПО      (на      момент      обнаружуния
отказа/аварии)'].fillna(dataset['Кол-во      СПО      (на      момент
обнаружуния отказа/аварии)'].mode()[0], inplace=True)

dataset.describe()

# распределение значений
x_columns = dataset.columns[:-1]

for column in x_columns:
    dataset[column].hist()
    plt.title(column)
    plt.show()
```

```

# Корреляции.
# Большинство значений слабо коррелируют с целевой
переменной или друг с другом
plt.figure(figsize=(10, 10))
sns.heatmap(np.abs(dataset.corr()))

# зависимость целевой переменной от остальных переменных
for column in x_columns:
    dataset.plot(x=column, y='Наработка НКТ, сут', style='o',
alpha=0.2)
    plt.title(f'Зависимость целевой переменной от
"{column}")
    plt.xlabel(column)
    plt.ylabel('Наработка НКТ, сут')
    plt.show()

X = dataset.iloc[:, :-1].values
print(X)

Y = dataset.iloc[:, -1].values
print(Y)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, Y,
test_size=0.2, random_state=0)

# стандартизация входных переменных
mean = X_train.mean(axis=0)
std = X_train.std(axis=0)
X_train -= mean
X_train /= std
X_test -= mean
X_test /= std

# стандартизация выходных переменных, для меньших значений
функции ошибки (это может навредить тренировке сети)
y_mean = y_train.mean()
y_std = y_train.std()
y_train_scaled = (y_train - y_mean) / y_std
y_test_scaled = (y_test - y_mean) / y_std

from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train_scaled)

print(regressor.intercept_)

print(regressor.coef_)

y_pred = regressor.predict(X_test)

from sklearn import metrics

```

```

#  $\phi$ 
def display_regression_results(predicted_scaled_values):
    predicted_values = predicted_scaled_values * y_std +
y_mean
    df = pd.DataFrame({'Actual': y_test, 'Predicted':
predicted_values})
    display(df)

    plt.figure(figsize=(8, 8))
    plt.title('True and predicted values')
    plt.scatter(y_test, predicted_values)
    plt.xlabel('True value')
    plt.ylabel('Predicted value')
    plt.plot([0, 2000], [0, 2000], color='red')
    plt.show()

    print('Mean Absolute Error:',
metrics.mean_absolute_error(y_test, predicted_values))
    print('Mean Squared Error:',
metrics.mean_squared_error(y_test, predicted_values))
    print('Root Mean Squared Error:',
np.sqrt(metrics.mean_squared_error(y_test, predicted_values)))

    display_regression_results(y_pred)

from keras.models import Sequential
from keras.layers import Dense, Dropout
from tensorflow.keras import regularizers

def plot_training(history):
    train_loss = history.history['mae']
    val_loss = history.history['val_mae']
    epochs = len(train_loss)

    plt.figure(figsize=(8, 8))
    plt.title('Training and validation loss')
    plt.plot(range(1, epochs+1), [a*2 for a in train_loss],
label='Training MAE')
    plt.plot(range(1, epochs+1), val_loss, label='Validation
MAE')
    plt.xlabel('Epochs')
    plt.ylabel('MAE')
    plt.legend()
    plt.grid()
    plt.show()

model = Sequential([
    Dense(8, activation='elu'),
    Dense(1)
])
model.compile(optimizer='adam', loss='mse',
metrics=['mae'])

```

```

    history = model.fit(X_train, y_train_scaled, epochs=100,
batch_size=32,          verbose=2,          validation_data=(X_test,
y_test_scaled))

    plot_training(history)

    display_regression_results(model.predict(X_test)[: , 0])

    model = Sequential([
        Dense(16,          activation='relu',
kernel_regularizer=regularizers.L1()),
        Dense(16,          activation='relu',
kernel_regularizer=regularizers.L1()),
        Dense(16,          activation='relu',
kernel_regularizer=regularizers.L1()),
        Dense(1)
    ])
    model.compile(optimizer='adam',          loss='mse',
metrics=['mae'])
    history = model.fit(X_train, y_train_scaled, epochs=200,
batch_size=64,          verbose=2,          validation_data=(X_test,
y_test_scaled))

    plot_training(history)

    display_regression_results(model.predict(X_test)[: , 0])

```