

Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.04.01 «Информатика и вычислительная техника»  
 Отделение информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Анализ популярности блогера у его целевой аудитории в социальных сетях

Студент

Группа	ФИО	Подпись	Дата
8BM72	Погожев Александр Олегович		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Губин Евгений Иванович	к.т.н., доцент		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ШБИП	Меньшикова Екатерина Валентиновна	к.ф.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООД	Алексеев Николай Архипович			

### ДОПУСТИТЬ К ЗАЩИТЕ:

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

## Планируемые результаты обучения

Код	Результаты обучения	Требования ФГОС ВО, СУОС, критерии АИОР, требования профессиональных стандартов
<b>Общепрофессиональные компетенции</b>		
P1	Самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.	Требования ФГОС ВО, СУОС ТПУ (УК-1,4, 5; ОПК-1,2, ПК-2,7,12,14,15), критерий 5 АИОР (п. 1.1), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P2	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.	Требования ФГОС ВО, СУОС ТПУ (УК-1,4, ОПК-1,5, ПК-7-9), критерий 5 АИОР (п. 1.1, 1.2), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.	Требования ФГОС ВО, СУОС ТПУ (УК-1,6, ОПК-1,2,5,6, ПК-1,7), критерий 5 АИОР (п. 1.2), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P4	<p>Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Владеть, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, применять специальную лексику и профессиональную терминологию языка.</p> <p>Демонстрировать способность к самостоятельному обучению новым методам исследования, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения.</p>	Требования ФГОС ВО, СУОС ТПУ (УК-4,6, ОПК-3,4), критерий 5 АИОР (п. 1.6, п. 2.2,2.6.), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.

P5	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современных информационных систем, в управлении коллективом. Способность организовывать и руководить работой команды.	Требования ФГОС ВО, СУОС ТПУ (УК-2,3,6, ОПК-1,6, ПК-4,5,7,8), критерий 5 АИОР (п. 2.1, п. 2.3, п. 1.5), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P6	Совершенствовать и развивать свой интеллектуальный и общекультурный уровень. Проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности.	Требования ФГОС ВО, СУОС ТПУ (УК-1,6, ОПК-3,5, ПК-7,8), критерий 5 АИОР (п. 2.4, п. 2.5) , соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
<b>Профиль 1 «Профиль «Сети ЭВМ и телекоммуникаций»</b>		
P7	Разрабатывать стратегии и цели проектирования, критерии эффективности и ограничения применимости, новые методы, средства и технологии проектирования сетей ЭВМ и телекоммуникаций.	Требования ФГОС ВО, СУОС ТПУ (УК-1,3,4,5, ПК-1,3,7,8,9,14, ОПК-1,2,5,6.), критерий 5 АИОР (п.1.3), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей. Требования профессиональных стандартов 06.001, 06.015, 40.057, 06.011, 06.016.
P8	Осуществлять авторское сопровождение процессов проектирования, внедрения, эксплуатации и модернизации сетей ЭВМ и телекоммуникаций на всех этапах жизненного цикла.	Требования ФГОС ВО, СУОС ТПУ (УК-2-4, ПК-4,16,17, ОПК-1,2,6.), критерий 5 АИОР (п. 1.5), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей. Требования профессиональных стандартов 06.001, 06.015, 40.057, 06.011, 06.016.

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

Школа ИШИТР  
 Направление подготовки 09.04.01 Информатика и вычислительная техника  
 Отделение информационных технологий

**УТВЕРЖДАЮ:**  
 Руководитель ООП  
 \_\_\_\_\_ И.А. Ботыгин  
 (Подпись)    \_\_\_\_\_ (Дата)    (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

Магистерской диссертации
--------------------------

Студенту:

Группа	ФИО
8ВМ72	Погожеву Александру Олеговичу

Тема работы:

Анализ популярности блогера у его целевой аудитории в социальных сетях	
Утверждена приказом директора (дата, номер)	№ 2764/с от 09.04.2019

Срок сдачи студентом выполненной работы:	
--	--

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b>  <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><i>В поставляемую систему транскодирования каналов головной станции необходимо добавить возможность записи контента на удаленный сервер.</i></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. Аналитический обзор протоколов обмена данными</li> <li>2. Объект и методы исследования</li> <li>3. Реализация серверной и клиентской сторон</li> </ol>

<b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i>		
<b>Консультанты по разделам выпускной квалификационной работы</b> <i>(с указанием разделов)</i>		
<b>Раздел</b>	<b>Консультант</b>	
<b>Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</b>	Доцент ОСГН, к.ф.н., Меньшикова Е.В.	
<b>Социальная ответственность</b>	Ассистент ООД, Алексеев Н.А	
<b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b>		

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

**Задание выдал руководитель:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИТ ИШИТР	Губин Евгений Иванович	к.т.н., доцент		

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8ВМ72	Погожев Александр Олегович		

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

Студенту:

Группа	ФИО
8BM72	Погожев Александр Олегович

Школа	Институт кибернетики	Отделение школы (НОЦ)	Информационных систем и технологий
Уровень образования	Магистр	Направление/специальность	Сети ЭВМ и телекоммуникации

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

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Бюджет = 191 512,23руб. Зарплата = 92 693,55 руб. Доп. зарплата = 9 269,36 руб. Накладные расходы = 30 588,87 руб. Внебюджетные отчисления = 25 119,95 руб. Стоимость оборудования = 29900 руб. Стоимость материалов = 566,5 руб. Стоимость работы сторонних организаций = 3363,56
2. Нормы и нормативы расходования ресурсов	Работа с нормативно-правовыми документами РФ.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Положения ФЗ РФ №212 от 24 июля 2009 г. «О страховых взносах в пенсионный фонд Российской Федерации, фонд социального страхования Российской Федерации, Федеральный фонд обязательного медицинского страхования»; Общая система налогообложения с учетом льгот для образовательных учреждений.

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

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИТ с позиции ресурсоэффективности и ресурсосбережения	Оценка потенциальных потребителей результатов исследования; анализ конкурентных технических решений. Определение возможных альтернатив с помощью морфологического анализа.
2. Планирование и формирование бюджета научных исследований	Расчет материальных затрат НИТ; Расчет затрат на специальное оборудование для научных работ; Основная заработная плата исполнителей темы; Дополнительная заработная плата исполнителей темы; Отчисление во внебюджетные фонды; Прочие расходы.
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	Анализ интегральных показателей эффективности.

**Перечень графического материала (в том числе таблиц):**

1. «Портрет» потребителя результатов НТИ
2. Сегментирование рынка
3. Оценка конкурентоспособности технических решений
4. График проведения и бюджет НТИ
5. Оценка ресурсной, финансовой и экономической эффективности НТИ
6. Потенциальные риски

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

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отделения социально- гуманитарных наук	Меньшикова Екатерина Валентиновна	к.ф.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ72	Погожев Александр Олегович		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ВМ72	Погожев Александр Олегович

Школа	ИШИТР	Отделение школы (НОЦ)	Информатика и вычислительная техника
Уровень образования	Магистр	Направление/специальность	Сети ЭВМ и телекоммуникации

### Исходные данные к разделу «Социальная ответственность»:

<p><i>Описание рабочего места (рабочей зоны, технологического процесса, механического оборудования) на предмет возникновения:</i></p> <p><i>вредных проявлений факторов производственной среды (метеоусловия, вредные вещества, освещение, шумы, вибрации, электромагнитные поля, ионизирующие излучения)</i></p> <p><i>опасных проявлений факторов производственной среды (механической природы, термического характера, электрической, пожарной и взрывной природы)</i></p> <p><i>негативного воздействия на окружающую природную среду (атмосферу, гидросферу, литосферу)</i></p> <p><i>чрезвычайных ситуаций (техногенного, стихийного, экологического и социального характера)</i></p>	<p><i>Анализ и выявление вредных производственных факторов рабочей среды, а именно: электромагнитное излучение, микроклимат, освещение, шумы и прочие, влияющие на организм человека при разработке программного обеспечения в помещении учебной аудитории.</i></p> <p><i>Анализ и выявление опасных производственных факторов проектируемой среды, а именно: электробезопасность и пожаробезопасность.</i></p> <p><i>Утилизация люминесцентных ламп – основной источник загрязнения литосферы.</i></p> <p><i>Чрезвычайная ситуация техногенного характера для данного помещения – пожар.</i></p> <p><i>В качестве исходных данных использованы параметры рабочего помещения, в котором производилась разработка и условия труда при работе с персональным компьютером.</i></p>
<p><i>Перечень законодательных и нормативных документов по теме</i></p>	<p><i>СанПиН 2.2.2/2.4.1340-03, СНИП 23-05-95, ГОСТ 6825-91, ГОСТ 12.1.003-83, СНИП 23-03-2003, СанПиН 2.2.4.548-96, ГОСТ 12.0.005-74, ГОСТ Р 51768-2001, 51057-01, ГОСТ 12.10.019</i></p>

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

<p><i>Анализ выявленных вредных факторов проектируемой производственной среды в следующей последовательности:</i></p> <p><i>физико-химическая природа вредности, её связь с разрабатываемой темой;</i></p> <p><i>действие фактора на организм человека;</i></p> <p><i>приведение допустимых норм с необходимой размерностью (со ссылкой на соответствующий нормативно-технический документ);</i></p> <p><i>предлагаемые средства защиты (сначала коллективной защиты, затем – индивидуальные защитные средства)</i></p>	<p><i>Анализ выявленных вредных факторов труда разработчика-программиста:</i></p> <p><i>недостаточная освещенность рабочей зоны;</i></p> <p><i>отклонение параметров микроклимата в помещении;</i></p> <p><i>повышенный уровень шума;</i></p> <p><i>повышенный уровень излучения электромагнитных полей.</i></p>
---	--

<p><i>Анализ выявленных опасных факторов проектируемой производственной среды в следующей последовательности механические опасности (источники, средства защиты); термические опасности (источники, средства защиты); электробезопасность (в т.ч. статическое электричество, молниезащита – источники, средства защиты); пожаровзрывобезопасность (причины, профилактические мероприятия, первичные средства пожаротушения)</i></p>	<p><i>Анализ выявленных опасных производственных факторов рабочей среды, влияющих на организм человека при разработке программного обеспечения в рабочем помещении учебной аудитории, а именно: опасность поражения электрическим током, опасность поражения статическим электричеством и пожароопасность.</i></p>
<p><i>Охрана окружающей среды: защита селитебной зоны анализ воздействия объекта на атмосферу (выбросы); анализ воздействия объекта на гидросферу (сбросы); анализ воздействия объекта на литосферу (отходы); разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды.</i></p>	<p><i>Утилизация используемой орг.техники и люминесцентных ламп.</i></p>
<p><i>Защита в чрезвычайных ситуациях: перечень возможных ЧС на объекте; выбор наиболее типичной ЧС; разработка превентивных мер по предупреждению ЧС; разработка мер по повышению устойчивости объекта к данной ЧС; разработка действий в результате возникшей ЧС и мер по ликвидации её последствий</i></p>	<p><i>Чрезвычайная ситуация техногенного характера для данного помещения – пожар. Установка общих правил поведения и рекомендаций во время пожара, план эвакуации.</i></p>
<p><i>Правовые и организационные вопросы обеспечения безопасности: специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства; организационные мероприятия при компоновке рабочей зоны</i></p>	<p><i>Основные проводимые правовые и организационные мероприятия по обеспечению безопасности трудящихся в учебных аудиториях.</i></p>
<p><b>Перечень графического материала:</b></p>	
<p><i>При необходимости представить эскизные графические материалы к расчётному заданию (обязательно для специалистов и магистров)</i></p>	<p><i>План эвакуации при пожаре. Схема организации рабочего места.</i></p>

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

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Алексеев Николай Архипович			

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ72	Погожев Александр Олегович		

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

Школа ИШИТР

Направление подготовки 09.04.01 Информатика и вычислительная техника

Уровень образования – Магистр

Отделение информационных технологий

Период выполнения – весенний семестр 2018 /2019 учебного года

Форма представления работы:

Магистерская диссертация

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

Срок сдачи студентом выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
28.05.19	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.	10
03.06.19	Социальная ответственность	10
05.06.19	Основная часть	70
07.06.19	Английская часть	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Губин Евгений Иванович	к.т.н., доцент		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Ботыгин И.А.	к.т.н., доцент		

## РЕФЕРАТ

Выпускная квалификационная работа содержит 120 страниц, 30 рисунков, 10 таблиц, 35 литературных источников, 1 приложение.

Ключевые слова: регрессия, нейронные сети, TensorFlow, fuzzy logic, decision tree, Python, комплаентность аудитории, анализ данных.

Предметом исследования является изучение основных принципов анализа данных, построения алгоритмов и моделей в задаче регрессии, выбор наилучшей конфигурации системы.

Объектом исследования является набор данных по активностям на страницах блогеров.

Цель работы – разработать новый подход для определения степени воздействия блогера на целевую аудиторию.

В процессе исследования проводился анализ предметной области, выбор программной части системы, ее организация и реализация, анализ данных, их предобработка, формирование моделей и их обучение.

Основными технологическими особенностями является использование языка Python 3.7, для анализа данных использовались библиотеки NumPy и Pandas, для построения моделей использовались библиотеки TensorFlow.

Область применения: рекламодатели и рекламные агентства.

Экономическая значимость работы представляет собой получение прогноза данных в автоматическом режиме для дальнейшей работы системы, где применяется разработанная программа.

В будущем планируется расширить функционал разработанной программы и увеличить точность прогнозов.

## ОГЛАВЛЕНИЕ

РЕФЕРАТ .....	11
1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ .....	16
1.1 Выбор программных компонентов .....	16
1.2 Разведочный анализ данных .....	18
1.2.2 Параметры с низкой дисперсией.....	19
1.2.3 Корреляция параметров.....	20
1.2.4 Определение значимости параметров .....	20
1.3 Разложение на главные компоненты .....	21
1.4 Подготовка данных.....	22
1.5 Нейронные сети .....	24
1.5.1 Функции активации .....	26
1.6 Аналитический обзор социальных сетей .....	27
1.6.1 Аудитории социальных сетей .....	29
1.7 Аналитический обзор инструментов сегментации аудитории.....	33
1.8 Оценка влияния блогера на аудиторию.....	35
2 АНАЛИЗ ДАННЫХ .....	37
2.1 Предобработка .....	37
2.2 Оценка комплаентности аудитории.....	39
3 ФОРМИРОВАНИЕ МОДЕЛЕЙ И РЕЗУЛЬТАТ РАБОТЫ .....	41
3.1 Нейронная сеть .....	41
3.2 Нечёткая логика .....	45
4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ,РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ .....	48
5 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ .....	66
ЗАКЛЮЧЕНИЕ .....	86

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ ..... 87

## ВВЕДЕНИЕ

Блоггинг становится все более популярным и удобным путем для общения, публикации информации, обсуждения интересов, высказывания мнений и предположений, распространения новостей и формирования виртуальных сообществ. Блогосфера подчиняется распределению по степенному закону: очень мало блогов чрезвычайно влиятельны, а огромное количество блогов практически неизвестно. Независимо от того, является ли (мультиавторский) блог влиятельным или нет, есть влиятельные блогеры. Однако огромное количество таких блогов делает чрезвычайно трудным изучение каждого из них. Один из способов проанализировать эти блоги - найти влиятельных блогеров и считать их представителями сообщества. Влиятельные блогеры могут воздействовать на других блогеров различными способами.

Развитие веб-приложений с широким участием (или Web 2.0 (O'Reilly2005)) создало онлайн-медиа, которые превращают бывших потребителей массовой информации в настоящих производителей информации (Gillmor2006). Примерами могут служить блоги, вики, совместные теги, совместное использование мультимедиа и другие подобные услуги. Сайт блога или просто блог (сокращение от веб-журнала) - это набор записей отдельных лиц. Эти записи, называемые постами, могут объединять текст, изображения и ссылки на другие блоги, посты блога и / или веб-страницы. Ведение блога становится популярным средством для пользователей Интернета выразить мнение, общаться, делиться, сотрудничать, обсуждать и размышлять. Блогеры, авторы блогов, свободно формируют свои сообщества, представляющие особый интерес, где они делятся мыслями, выражениями, обсуждают идеи и предлагают предложения в интерактивном режиме. Блогосфера предоставляет платформу, способствующую созданию виртуальных сообществ с особыми интересами. Блогосфера и социальные сети в целом изменяют бизнес-модели, способствуют развитию вирусного маркетинга, предоставляют возможности для анализа тенденций и прогнозирования продаж, помогают в борьбе с терроризмом и выступают в качестве источника массовой информации.

В физическом мире, согласно Келлеру и Берри (2003), 83% людей предпочитают консультироваться с семьей, друзьями или экспертами по сравнению с традиционной рекламой, прежде чем попробовать новый ресторан, 71% людей делают то же самое, прежде чем покупать лекарства по рецепту или посещать развлекательные места, и 61% людей перед просмотром фильма поговорят о нем с семьей или друзьями. Короче говоря, прежде чем люди покупают или принимают решения, они слушают чужой опыт, мнения и предположения. Люди, чей опыт, мнения и предложения востребованы, называются «влиятельными лицами». Проводя параллели между физическими и виртуальными сообществами, среди граждан блогосферы, мы хотим выяснить, существуют ли факторы влияния в виртуальном сообществе (блоге), что это за факторы, как их определить и как на них повлиять.

Дальше в вкр будет предложен коэффициент влияния блогера на аудиторию для дальнейшего использования в дальнейшем прогнозировании эффективности рекламы или поста на аудиторию. Для этого будут применены методы работы с данными, нейронные сети и не чёткая логика, для поиска не интуитивных параметров и нахождение зависимостей.

# 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

## 1.1 Выбор программных компонентов

Машинное обучение является быстроразвивающейся и востребованной сферой информационных технологий. Множество корпораций и исследовательских центров разрабатывают алгоритмы и программные компоненты, и инструменты, решающие типовые задачи анализа данных. Наиболее популярные из доступных инструментов с открытым исходным кодом являются библиотеки, основанные на языках C++ и Python [2].

Согласно данным анализа рынка, наиболее популярным решением является использование Python+Tensorflow. Когда TensorFlow был с открытым исходным кодом в ноябре 2015 года, уже было много популярных библиотек с открытым исходным кодом для глубокого обучения и большинство функций TensorFlow уже существовали в той или иной библиотеке. Краткий анализ возможных решений представлен в таблице ниже [3].

Таблица 1 – Популярные библиотеки машинного обучения

Библиотека	API	Платформы	Основатель	Год
Caffe	Python, C++, Matlab	Linux, macOS, Windows	Y. Jia, UC Berkeley (BVLC)	2013
Torch	C++, Lua	Linux, macOS, iOS, Android	R. Collobert, K. Kavukcuoglu, C. Farabet	2002
H2O	Python, R	Linux, macOS, Windows	H2O.ai	2014
MXNet	Python, C++, др.	Linux, macOS, Windows, iOS, Android	DMLC	2015
TensorFlow	Python, C++, JS, др.	Linux, macOS, Windows, iOS,	Google	2015
Theano	Python	Linux, macOS, iOS	University of Montreal	2010

Тем не менее, чистый дизайн TensorFlow, масштабируемость, гибкость, и отличная документация (при поддержке Google) быстро поднял его в начало списка. TensorFlow был спроектирован так, чтобы быть гибким, масштабируемым и готовым к работе, а существующие структуры, возможно, реализуют только две из трех особенностей. Некоторые из основных моментов TensorFlow:

- Он работает не только на Windows, Linux и MacOS, но и на мобильных устройствах, включая как iOS, так и Android.
- Он предоставляет очень простой Python API под названием TF.Learn ([tenorflow.contrib.learn](https://www.tensorflow.org/contrib/learn)), совместимый с Scikit-Learn. Он может быть использован для обучения различных типов нейронных сетей всего в несколько строк кода.
- Несколько других высокоуровневых API были построены независимо от TensorFlow, такие как Keras (теперь доступно в [tenorflow.contrib.keras](https://www.tensorflow.org/contrib/keras)) или Pretty Tensor.
- Его основной Python API предлагает гораздо большую гибкость (за счет более высокой сложности) для создания всевозможных вычислений, включая любую архитектуру нейронной сети, которую возможно реализовать.
- Он включает в себя высокоэффективные реализации C++ многих операций ML, в частности те, которые необходимы для построения нейронных сетей. Существует также C++ API для определения собственных высокопроизводительных операций.
- Он предоставляет несколько продвинутых узлов оптимизации для поиска параметров, которые минимизируют функцию потерь. Они очень просты в использовании, так как TensorFlow автоматически заботится о вычислении градиентов функций, которые определяет разработчик. Это называется автоматическое дифференцирование.

## 1.2 Разведочный анализ данных

В статистике разведочный анализ данных (РДА) — это подход к анализу наборов данных для обобщения их основных характеристик, часто с помощью визуальных методов. Статистическая модель может использоваться или может быть отвергнута, но в первую очередь РДА предназначен для того, чтобы увидеть, что данные могут проявить, помимо формальной задачи моделирования или проверки гипотез. Джон Тьюки продвигал исследовательский анализ данных, чтобы побудить статистиков исследовать данные и, возможно, сформулировать гипотезы, которые могли бы привести к сбору новых данных и экспериментам [4].

Существует ряд инструментов, которые полезны для РДА, типичные графические методы:

- Диаграмма размаха
- Гистограмма
- Многовариантная диаграмма
- Точечный график
- Стебель-лист диаграмма
- Параллельные координаты
- Целевая проекция преследования
- Уменьшение размерности:
  - Многомерное масштабирование
  - Анализ главных компонент (РСА)
  - Полилинейный РСА
  - Нелинейное уменьшение размерности (NLDR)
- Методы проецирования

### **1.2.1 Отсутствие значений параметра**

Пропущенные значения отражают беспорядок в данных реального мира. Причин их возникновения может быть множество - от человеческих ошибок при вводе данных, неправильных показаний датчиков до программных ошибок в конвейере обработки данных.

Наиболее простой метод обработки таких данных - удаление. Это быстрое решение может работать в некоторых случаях, когда доля пропущенных значений является относительно низкой (<10%), в большинстве случаев это приведет к потере большого объема полезных данных [5].

В более общем случае, стандартный и часто очень подходящий подход заключается в замене пропущенных вещественных чисел средним, медианным или модальным значениями.

При значениях параметра в виде категории, стандартная процедура, которую нужно сделать, это заменить пропущенную запись самой часто встречающейся.

### **1.2.2 Параметры с низкой дисперсией**

Интуитивно можно предположить, что функции с малой дисперсией бесполезны и являются просто шумом для модели [6]. Тем не менее, есть некоторые моменты, которые необходимо учитывать при работе с подобными данными:

- Дисперсия объекта не является единичной, и, повторно выражая, скажем, длину в метрах, миллиметрах или футах, изменяется дисперсия.
- Дисперсия функции не учитывает связь между функцией и ответом системы, которая находится в центре внимания моделей обучения с учителем. В то время как предиктор может иметь небольшие отклонения, связь между предиктором и ответом может быть очень сильной в этом диапазоне.

### **1.2.3 Корреляция параметров**

Основной целью регрессионного анализа является изоляция взаимосвязи между каждой независимой переменной и зависимой переменной. Интерпретация коэффициента регрессии заключается в том, что он представляет среднее изменение зависимой переменной для каждого изменения на 1 единицу в независимой переменной, когда все остальные независимые переменные остаются постоянными. Эта последняя часть имеет решающее значение для мультиколлинеарности [7].

Мультиколлинеарность – наличие линейной зависимости между объясняющими переменными (факторами) регрессионной модели. Когда независимые переменные коррелируют, это указывает на то, что изменения в одной переменной связаны со сдвигами в другой переменной. Чем сильнее корреляция, тем сложнее изменить одну переменную без изменения другой. Для модели становится трудно оценить взаимосвязь между каждой независимой переменной и зависимой переменной независимо, потому что независимые переменные имеют тенденцию изменяться одновременно.

Одним их эффективных методов является удаление высококоррелированных предикторов из модели. Поскольку они предоставляют избыточную информацию, удаление одного из коррелированных факторов обычно не приводит к значительному снижению  $R^2$ .

### **1.2.4 Определение значимости параметров**

В приложениях интеллектуального анализа данных входные предикторы редко бывают одинаково значимы. Часто лишь немногие из них оказывают существенное влияние на ответ; подавляющее большинство не имеет значения и могут быть исключены. Часто полезно узнать относительную важность или вклад каждой входной переменной в прогнозирование ответа [8].

Преимущество использования ансамблей методов деревьев решений, таких как градиентный бустинг, заключается в том, что они могут автоматически предоставлять оценки важности параметров из обученной прогностической модели.

Как правило, важность обеспечивает оценку, которая указывает, насколько полезной был каждый параметр при построении расширенных деревьев решений в модели. Чем больше атрибут используется для принятия ключевых решений с деревьями решений, тем выше его относительная важность.

Эта важность вычисляется явно для каждого атрибута в наборе данных, что позволяет ранжировать атрибуты и сравнивать их друг с другом [9].

Для одного дерева решений  $T$ , Breiman et al. (1984) предложена в качестве меры релевантности для каждой переменной предиктора  $X_\ell$  следующая формула:

$$L_i^2(T) = \sum_{t=1}^{J-1} \hat{i}_t^2 I(v(t) = l) \quad (1)$$

Сумма вычисляется по  $J - 1$  внутренним узлам дерева. В каждом таком узле  $t$  одна из входных переменных  $X_{v(t)}$  используется для разделения области, связанной с этим узлом, на два субрегиона; внутри каждой отдельной области соответствуют значениям ответа. Выбранная переменная является той, которая дает максимальную оценку улучшение  $\hat{i}_t^2$  по сравнению с другими. Квадрат относительной важности переменной  $X_\ell$  является суммой таких улучшений в квадрате по всем внутренним узлам, для которых он был выбран в качестве переменной разделения. Эта мера важности легко обобщается на аддитивные разложения деревьев:

$$L_i^2 = \frac{1}{M} \sum_{m=1}^M L_i^2(T_m) \quad (2)$$

### 1.3 Разложение на главные компоненты

Многофакторный анализ часто работает с данными, включающими значительное количество коррелированных переменных. Анализ главных компонент (РСА) является инструментом уменьшения размеров, который можно использовать для редуцирования большого набора переменных в небольшой объем параметров, который по-прежнему содержит большую часть информации [10].

Анализ главных компонент является математической процедурой, которая преобразует ряд (возможно) коррелированных переменных в (меньшее) число некоррелированных переменных, называемые главными компонентами.

Основные этапы алгоритма следующие:

1. Вычислить ковариационную матрицу  $X$  точек данных.
2. Вычислить собственные векторы и соответствующие собственные значения.
3. Сортировать собственные векторы по их собственным значениям в порядке убывания.
4. Выбрать первые  $k$  собственных векторов, это будут новые  $k$  измерений.
5. Преобразовать исходные точек  $n$  измерений в  $k$  измерений.

В качестве оператора преобразования используется матрица собственных векторов, отсортированных по собственным значениям.

#### **1.4 Подготовка данных**

Перед использованием алгоритмов data mining необходимо произвести подготовку набора анализируемых данных. Так как ИАД может обнаружить только присутствующие в данных закономерности, исходные данные с одной стороны должны иметь достаточный объём, чтобы эти закономерности в них присутствовали, а с другой — быть достаточно компактными, чтобы анализ занял приемлемое время. Чаще всего в качестве исходных данных выступают хранилища или витрины данных. Подготовка необходима для анализа многомерных данных до кластеризации или интеллектуального анализа данных.

Далее данные фильтруются. Фильтрация удаляет выборки с шумами и пропущенными данными.

Отфильтрованные данные сводятся к наборам признаков (или векторам, если алгоритм может работать только с векторами фиксированной размерности), один набор признаков на наблюдение. Набор признаков формируется в соответствии с гипотезами о том, какие признаки сырых данных имеют высокую

прогнозную силу в расчете на требуемую вычислительную мощность для обработки. Например, черно-белое изображение лица размером 100×100 пикселей содержит 10 тыс. бит сырых данных. Они могут быть преобразованы в вектор признаков путём обнаружения в изображении глаз и рта. В итоге происходит уменьшение объёма данных с 10 тыс. бит до списка кодов положения, значительно уменьшая объём анализируемых данных, а значит и время анализа.

Ряд алгоритмов умеют обрабатывать пропущенные данные, имеющие прогностическую силу (например, отсутствие у клиента покупок определенного вида). Скажем, при использовании метода ассоциативных правил (англ.)русск. обрабатываются не векторы признаков, а наборы переменной размерности.

Выбор целевой функции будет зависеть от того, что является целью анализа; выбор «правильной» функции имеет основополагающее значение для успешного интеллектуального анализа данных.

Наблюдения делятся на две категории — обучающий набор и тестовый набор. Обучающий набор используется для «обучения» алгоритма data mining, а тестовый набор — для проверки найденных закономерностей. SEMMA

SAMPLE-определение входных данных, в том числе выборку (sample) из больших данных и разделение ее на тренировочную, валидационную и тестовую.

EXPLORE-исследование набора данных статистически и графически (графики, описательные статистики, определение важных переменных, ассоциативный анализ)

MODIFY- подготовка данных для анализа (создание дополнительных переменных или изменение существующих переменных для анализа, выявление выбросов ("outliers"), оценка отсутствующих данных ("missing"), изменение (модифицирование) способа использования входных переменных для анализа, проведение кластерного анализа, анализ с использованием сетей Кохрена и т.п.)

MODEL- оценка предсказательной модели (регрессионная модель, деревья решений, нейронные сети и др.)

ASSESS- сравнение конкурирующих предсказательных моделей (графическое сравнение исследуемых респондентов, графики доходности и т.п.)

### 1.5 Нейронные сети

Искусственные нейронные сети являются одним из основных инструментов, используемых в машинном обучении. В 1943 году МакКаллок и Питтс[11] предложили математическую модель нейрона. Это алгоритм, принимающий на входе  $n$ -мерный вектор признакового описания  $x = (x_1, \dots, x_n)$ . Поступающие в нейрон импульсы складываются с весами  $w_1, \dots, w_n$ . Если вес положительный, то соответствующий синапс возбуждающий, если отрицательный, то тормозящий. Если суммарный импульс превышает заданный порог активации  $w_0$ , то нейрон возбуждается и выдаёт на выходе 1, иначе выдаётся 0. Таким образом, нейрон вычисляет  $n$ -арную булеву функцию вида

$$a(x) = \varphi\left(\sum_{j=1}^n w_j x^j - w_0\right) \quad (3)$$

где  $\varphi(z)$  — функция активации.

Исходя из принципа минимизации эмпирического риска задача настройки синаптических весов может быть сведена к поиску вектора  $w$ , доставляющего минимум функционалу качества:

$$Q(w) = \sum_{i=1}^l L(a(x_i), y_i) \rightarrow \min_w \quad (4)$$

где  $L(a, y)$  — заданная функция потерь, характеризующая величину ошибки ответа  $a$  при правильном ответе  $y$ .

Используя подход градиентного спуска и функционал качества, можно записать уравнение итеративного обновления весов нейронной сети:

$$w = w - \eta \frac{dQ}{dw} \quad (5)$$

Искусственная нейронная сеть содержит множество параметров конфигурации, среди которых число нейронов  $M$  и скрытых слоев  $N$ . Для охвата большего количества значений используется дополнительный нейрон, называемый нейроном смещения [12].

Обучение таких сетей производится алгоритмом, основанным на градиентном спуске, называемым обратным распространением. Благодаря аналитическому дифференцированию суперпозиции с сохранением необходимых промежуточных величин, удается избежать большого количество операций, необходимых для вычисления градиента, возрастающих пропорционально числу весовых коэффициентов [13]. Он обладает рядом преимуществ:

- Достаточно высокая эффективность.
- Через каждый нейрон проходит информация только о связанных с ним нейронах.
- Высокая степень общности.

Таким образом, общий вид ИНС представлен на рисунке 1.

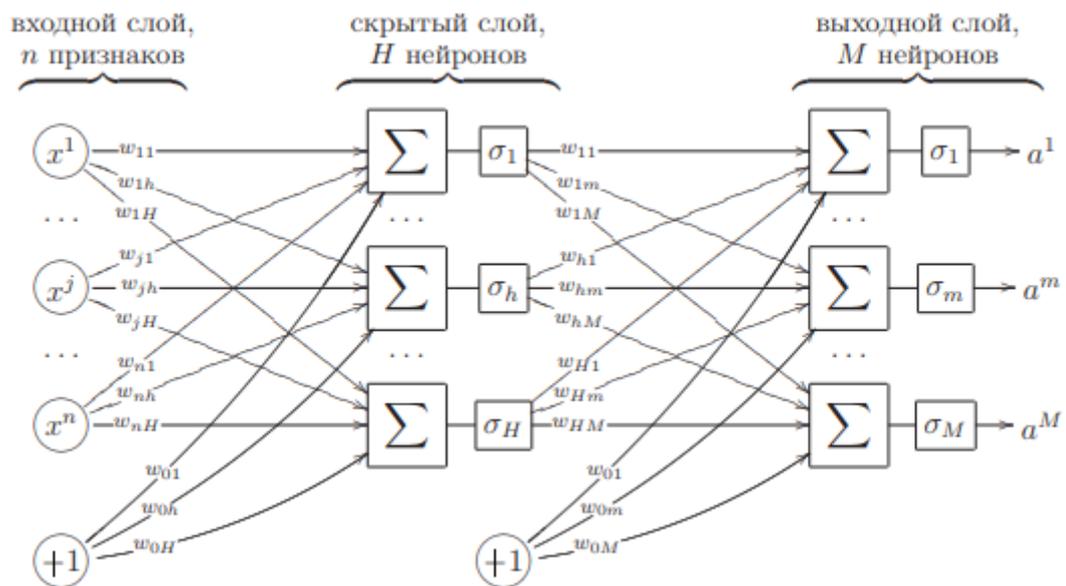


Рисунок 1 – Общий вид нейронной сети

### 1.5.1 Функции активации

Помимо этого каждый слой нейронной сети может содержать различные функции активации. Наиболее используемыми являются:

Сигмоидальная функция активации – это функция активации вида.

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (6)$$

Диапазон значений от 0 до 1. Это S-образная кривая. Ее легко анализировать и применить, но она обладает рядом недостатков, такие как затухание градиента и низкая скорость схождения.

Гиперболический тангенс – это функция активации вида:

$$f(x) = \frac{2}{(1 + e^{-2x})} - 1 \quad (7)$$

Она выражается через сигмоидальную функцию, следовательно имеет те же недостатки, за исключением того факта, что она ограничена диапазоном (-1, 1).

ReLU – функция дает вывод  $x$ , если  $x$  положительный, и 0 в противном случае.

$$f(x) = \begin{cases} x, & \text{если } x \geq 0 \\ 0 & \text{иначе} \end{cases} \quad (8)$$

Несмотря на то, что при такой активации есть шанс градиентного взрыва, она позволяет избежать затухания градиента. Из-за горизонтальной линии в ReLU (для отрицательного  $x$ ) градиент может идти к 0. Для активаций в этой области ReLU градиент будет равен 0, из-за чего веса не будут корректироваться во время градиентного спуска. Это означает, что те нейроны, которые входят в это состояние, перестанут реагировать на изменения в ошибке / вводе.

Экспоненциальная линейная единица - ELU - функция, которая, как правило, быстрее сводит потери к нулю и дает более точные результаты. Она имеет выход  $x$ , если  $x$  положительный и  $\alpha(\exp(x)-1)$  в противном случае.

$$f(x) = \begin{cases} x, & \text{если } x \geq 0 \\ \alpha(e^x - 1) & \text{иначе} \end{cases} \quad (9)$$

В отличие от других функций активации, ELU имеет дополнительную альфа-константу, которая должна быть положительным числом. ELU очень похож на ReLU, кроме отрицательных входов. Они оба находятся в форме идентичной функции для неотрицательных входных данных. С другой стороны, ELU медленно сглаживается до тех пор, пока его выход не станет равным  $-\alpha$ , тогда как ReLU резко сглаживается[14].

## 1.6 Аналитический обзор социальных сетей

«Социальная сеть» – сайт, при помощи которого можно общаться с другими пользователями ресурса. На таких интернет-площадках можно публиковать, просматривать и комментировать информацию других зарегистрированных участников: музыкальные, фото и видеоматериалы, документы, статьи.

Растущая популярность таких порталов побуждает использовать их не только для личных целей. Продвижение и поддержка компаний, продажи товаров, услуг – цели стратегий в социальном маркетинге (SMM). Ключ к успеху – выбор правильной сети среди множества имеющихся.

Социальные сети для общения (Relationship networks):

Это самые распространенные и востребованные социальные медиа на сегодняшний день. К ним относятся Facebook, Вконтакте, Одноклассники, LinkedIn. И хотя данный вид социальных медиа не появился первым, он стал определяющим для всей отрасли. Сегодня страница бренда в Twitter или Facebook - это общепринятый стандарт.

### Социальные сети для обмена медиа-контентом (Media sharing networks):

Данный вид социальных медиа дает пользователям широкие возможности для обмена видео- и фото-контентом. Сам принцип распространения информации и нативные возможности, например, фильтры в Instagram, дают таким социальным медиа преимущество перед многофункциональными собратьями. Отличительной особенностью является также и масштабирование контента: некоторые предлагают публиковать короткие видеоролики, другие дают вам возможность создать собственный видеоканал.

### Социальные сети для авторских записей (Social publishing platforms)

К данному типу социальных медиа относятся сервисы для блоггинга и микро-блоггинга, где пользователи создают и публикуют текстово-медийный контент. Сюда относятся такие популярные платформы, как Twitter, Medium и Tumblr.

### **Причины популярности**

Огромный интерес к сайтам для общения, который наблюдается последние годы, вполне объясним. Среди основных причин:

- 1) Коммуникация без границ, рамок и ограничений. Где и как еще можно подружиться с владельцами крупного бизнеса, знаменитостями, коллегами или людьми с похожими увлечениями.
- 2) Поиск и восстановление утерянных связей. Узнать о судьбе своих одноклассников, бывших коллег, поделиться своими успехами теперь очень просто.
- 3) Повышение самооценки – лайки и комплименты всегда ценятся. Даже если они виртуальные.
- 4) Коммерческие цели – продажа и покупка товаров и услуг через социальные сети давно стали нормой. Цены, доступней чем в магазинах, есть отзывы реальных покупателей, доверие благодаря

рекомендациям друзей. Для продавцов экономия на рекламе, персонале, аренде торгового места.

- 5) Удобство и простота использования – зайти в любимое приложение можно с мобильного телефона по дороге домой, в перерыве на работе, на перемене между парами.

В данной работе будут рассматриваться самые популярные в своих сегментах социальные сети: Facebook, Twitter, Instagram, LinkedIn, VKontakte, Youtube.

### **1.6.1 Аудитории социальных сетей**

#### **Тенденции и перспективы**

Отчет за 2018 год компаний We Are Social и Hootsuite показывает убедительный рост пользователей в социальных ресурсах. Зарегистрированы в тех или иных приложениях свыше 3 млрд человек. Ежедневно происходит 1 млн новых регистраций. Причем растет количество заходов на сайты с мобильных устройств – почти 90% посетителей используют смартфоны и планшеты, а не стационарные компьютеры.

В то же время, если пару лет назад на площадках для общения можно было быстро узнавать новости и события из жизни реальных людей, то сейчас ситуация в корне меняется. Все больше появляется рекламы, мотиваторов и другой общей информации. И это уже не такой тесный, близкий и привычный круг общения. Падает уровень доверия, а значит и продать что-то становится сложнее.

Появляются новые и новые стартапы, которые хоть и не сдвинут Facebook в ближайшее время с лидерской позиции, но привлекут часть аудитории. Планировать маркетинговую кампанию требуется внимательней и с предварительным анализом выбранного ресурса, товара и целевой аудитории.

Исследовательский ресурс Ebizmba опубликовал рейтинг популярности мировых социальных сетей по состоянию на февраль 2019 года. Первую тройку

без особых сюрпризов сформировали Facebook, Twitter и LinkedIn. В топ попали сразу две соцсети, созданные на постсоветском пространстве – российская “ВКонтакте” и латвийская Ask.fm.

Ниже – основные показатели самых популярных социальных сетей мира.

## **Facebook**

Безоговорочное первое место обеспечивают свыше 2,2 млрд аккаунтов в мире. Facebook работает с 2004 года, создан Марком Цукербергом. Простой и бесплатный сервис общения, который позволяет:

- 1) Зарегистрировать собственную страничку, где выкладываются фото и другие материалы.
- 2) Находить других пользователей, добавлять их в друзья, узнавать их новости.
- 3) Делиться информацией из сети или от других участников, выкладывать видео и фотоматериалы.
- 4) Создавать группы (открытые или закрытые).
- 5) Узнавать новости.
- 6) Комментировать и оценивать публикации.

Компания Facebook Inc. (владелец сервиса) постоянно расширяет возможности. Гиперлокальная реклама (в определенном радиусе от рекламодателя), Snapchat (исчезающие истории), GIF-анимация для комментирования, сохранение интересных материалов, мессенджер и другие функции удерживают интерес пользователей.

## **Twitter**

Начал работать в США с 2006 года. Возможности: ведение блога короткими сообщениями (максимум 280 символов). Интересная опция – «Moments» (сюжеты с развитием). О резонансных, популярных новостях можно узнавать, не подписываясь на других пользователей. Свои микроблоги в Twitter ведут люди, известные в самых разных сферах от шоу-бизнеса до спорта. Общее количество аккаунтов превышает 330 миллионов.

## **LinkedIn**

Сеть для профессионалов. На ее страничках общаются свыше 500 млн пользователей из 200 стран. Основной упор не на развлечения, юмор, новости и прочий привычный для других сервисов контент, а резюме, портфолио, перечень навыков (с подтверждением от реальных людей), вакансии. Спросить совета у коллег, обсуждать профессиональные вопросы, искать работу или сотрудников, получение рекомендаций – задачи LinkedIn. Проект запущен в 2003 году, в 2016 выкуплен компанией Microsoft.

## **Instagram**

Приложение для публикации снимков и небольших видеороликов (до 60 секунд), просмотра материалов других участников, с возможностью оценки и комментирования. С момента запуска (2010 год), Instagram пришелся по душе более чем 1,1 млрд зарегистрированных участников. В 2012 году выкуплен Facebook Inc. Популярные функции:

- 1) Ведение фотоблогов.
- 2) Обработка фотографий – эффекты, рамки, специальные режимы съемки, коллажи.
- 3) Создание своего профиля и ленты публикаций.
- 4) Просмотр друзей, оценка и комментирование их публикаций.
- 5) Кросс-постинг – одновременное размещение материала в нескольких соцсетях.
- 6) Поиск нужной информации по хештегам (специальным отметкам).

Instagram – пользуется особой популярностью у молодежи возрастом до 25 лет и активно развивается. Эта сеть одна из наиболее перспективных для бизнеса, продвижения товаров, раскрутки бренда и услуг.

## **ВКонтакте**

Российская сеть, созданная в 2006 года. Принадлежит компании Mail.Ru Group. Поддерживает свыше 90 языков, зарегистрировано свыше 460 млн пользователей. Наиболее популярна у русскоязычных пользователей – лидер всех соцсетей России по использованию.

Возможности:

- 1) Создание своей страницы.
- 2) Переписка с другими участниками проекта.
- 3) Пересылка картинок, аудио- и видеозаписей.
- 4) Совместные онлайн-игры.

Именно этот ресурс чаще всего выбирают для интернет-магазинов и продаж. Этому способствуют широкий выбор возможностей и большая русскоязычная аудитория.

## **Youtube**

Видеохостинг YouTube от Google – самая популярная видеосеть в мире. Считается и социальной сетью, так как позволяет пользователям регистрироваться, общаться, делиться своими роликами. Здесь можно найти информацию на любой вкус от развлечений до обучающих курсов. Основана в феврале 2005 года в США. На начало 2018 года насчитывает свыше 1,5 млрд активных пользователей, а результат в несколько миллионов просмотров популярных видео вполне достижим.

Основные возможности:

- 1) Выкладывать свои видеоролики для просмотра.
- 2) Поддерживаются разные форматы: панорамы, HD, Full HD, MPEG-4, AVC и другие.
- 3) Редактировать видео из браузера при помощи программ редакторов (добавлять титры, аудиодорожки, изображения, переходы).
- 4) Комментировать, оценивать чужие видео или комментарии.

5) Создавать групповые чаты.

6) Вести онлайн-трансляции.

Компания пытается контролировать содержимое ресурса, периодически удаляя ролики, которые нарушают авторское право, признаны экстремистскими. В некоторых странах ресурс блокируется, за публикацию негодных властям материалов.

### **1.7 Аналитический обзор инструментов сегментации аудитории**

Сегментирование целевой аудитории – это разделение аудитории на группы, где они объединены по признаку схожих потребностей (запросов). Сегментация целевой аудитории и рынка важны в интернет-рекламе и продвижении в социальных сетях. Инструменты, которые используют для сегментирования, делят аудиторию на группы и позволяют послать наиболее релевантное рекламное сообщение каждой конкретной группе, в зависимости от предпочтений пользователей, а не внутренних ощущений рекламодателя. При одинаковых затратах на рекламу эффективность сегментированной кампании будет выше.

Для получения полной картины групп покупателей делят на сегменты целевой аудитории:

*По социально-экономическим характеристикам.* Образование, доход, социальные блага. Показывает предпочтения покупателя и платежеспособность.

*По географическому признаку.* Разделение по странам, городам, районам, регионам, улицам. В зависимости от масштаба проекта.

*По демографическому признаку.* Сюда входят возраст, пол, семейное положение. От этого зависит покупательская активность – у каждой группы она разная. Подростки склонны к импульсивным покупкам, а зрелые покупатели больше сравнивают и подходят к приобретению обдуманно.

*По психографическому критерию.* Этот признак довольно размыт, но его тоже стоит учесть. Социальный статус, образ жизни, определенные моменты биографии.

И чем больше параметров применить, тем больше вероятность попасть правильным предложением в правильную аудиторию. А если компания умеет работать с BigData, то это прямой путь к успеху.

Для получения информации о целевой аудитории, используются следующие SMM инструменты:

1. Анкетирование - дает информацию о целевых переходах, возрасте, предпочтениях, устройствах и т.д.
2. Внутренние и внешние опросы.
3. Внутренние инструменты и статистика социальных сетей
4. Системы аналитики.

Такие ресурсы как Google.Analytics, Яндекс.Аудитория, Webmaster.Mail.ru, Wordstat позволяют составить более полную картину целевой аудитории исходя из поисковых запросов – определить, среди какой аудитории данный запрос наиболее популярен – по полу, возрасту и географии.

### **Таргетирование в VKontakte.**

Церебро Таргет – приложение для поиска лояльной аудитории, включающее более 100 инструментов для работы с VK. Более 80 тысяч маркетологов используют Церебро

TargetHunter – более 150 инструментов для SMM и арбитража трафика ВКонтакте.

### **Таргетирование в Facebook, Snapchat и Instagram**

Aitarget – комплекс услуг для повышения продаж через социальные сети. Аудит аккаунта, настройка таргетной рекламы, сегментирование аудитории.

## Многофункциональные маркетинговые приложения

Livedune – детальная аналитика аккаунтов во всех популярных социальных сетях, поиск блогеров, работа с аудиторией.

GetBlogger – это маркетплейс блогеров с большим набором инструментов по эффективному подбору и размещению рекламы, таргетированной на конкретную целевую аудиторию.

### 1.8 Оценка влияния блогера на аудиторию

При выборе блогера неопытные маркетологи ориентируются на количество подписчиков у блогера, однако этого недостаточно.

Для оценки влияния блогера на свою аудиторию нужно в первую очередь проанализировать новостную ленту этой аудитории. Лента новостей в социальных сетях работает по принципу «умной ленты». Это значит, что пользователи видят в своей ленте контент, который им больше всего подходит. Выводы о том, что интересно пользователю, а что нет, делаются на основе предыдущих действий: что лайкали, комментировали и сохраняли.

В связи с этим возник термин «вовлеченность». На вовлеченность влияют: количество подписчиков, лайки к посту, комментарии, сохранения. Т.е. количество подписчиков уже не является самым важным и определяющим. Сервисы по аналитике профиля в Инстаграм, такие как EpicDetect или LiveDune, используют коэффициент вовлеченности ER. Обычно он рассчитывается по такой формуле:

$$ER = (\text{лайки} + \text{комментарии}) / \text{подписчики} * 100\%$$

ER можно посчитать к каждому посту отдельно, но распространенная практика – учитывать показания 20 – 30 постов аккаунта. Накрутки подписчиков и лайки с введением алгоритма умной ленты потеряли первоначальный сакральный смысл. Теперь надо бороться не только за лайки, но и за

комментарии (которые тоже можно накрутить). А хороший вдумчивый комментарий дорогого стоит.

По мнению экспертов в области SMM, ER меньше 2% – повод бить тревогу и менять стратегию работы с аккаунтом. При таком показателе вовлеченности посты аккаунта практически не видны пользователями. Это все равно, что разговаривать самому с собой.

Тональность комментариев. Большое количество комментариев не означает лояльную аудиторию. Если большинство реакций негативные, такие же отзывы получит и реклама продукта. Едва ли рекламодатель стремится к такому результату.

Количество просмотров. На YouTube и в VK это главный показатель, он важнее количества подписчиков.

Также при оценке влияния блогера на аудиторию полезно посмотреть на его показатели в течение времени. Так резкий приток подписчиков за малый промежуток времени может свидетельствовать об использовании сервисов накрутки, следовательно большая часть аудитории блогера – боты; а отсутствие роста показателей в течение длительного времени говорит о том, что интерес к блогеру достиг предела и скоро пойдет на спад.

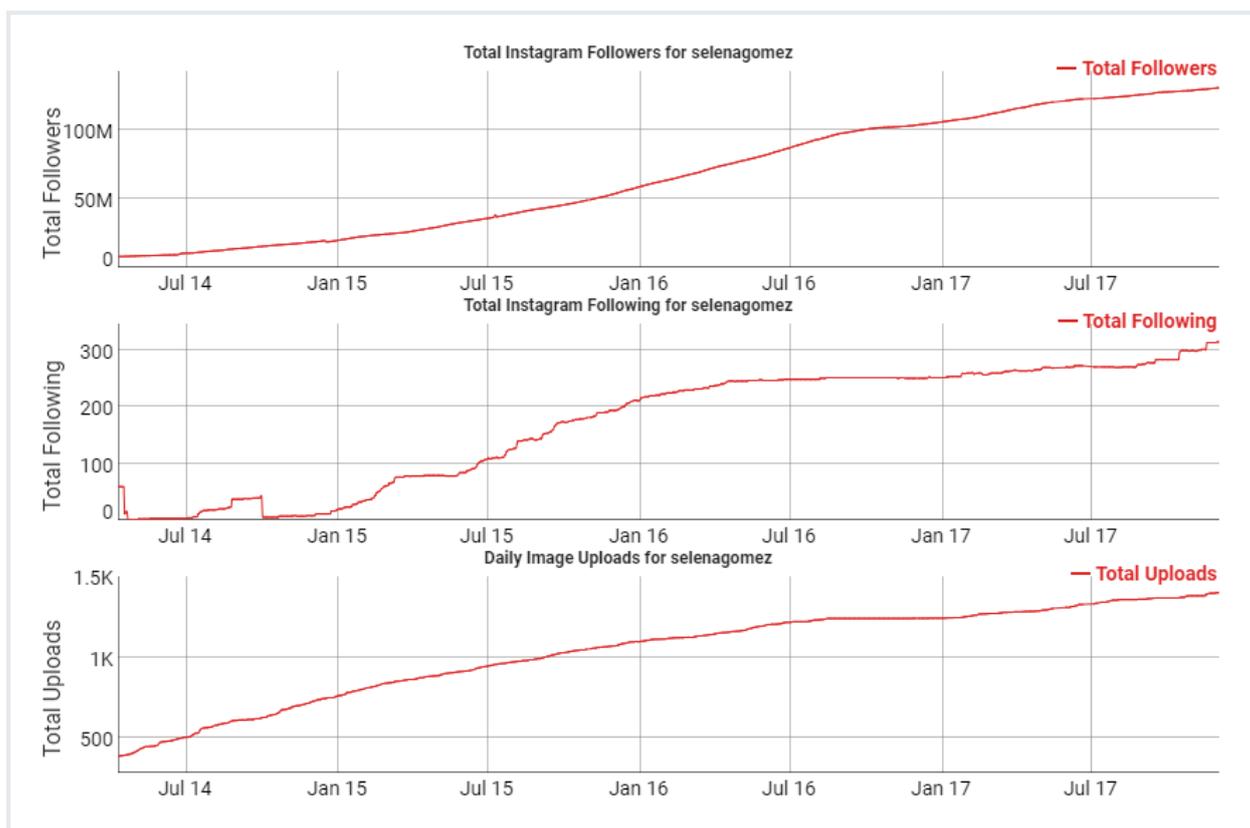


Рисунок 2 – динамика показателей аккаунта блогера за год.

Доскональное изучение аудитории блогера также очень важно. Нередко среди подписчиков встречаются всевозможные боты и бизнес-аккаунты, не представляющие никакой ценности при проведении рекламной компании. Для эффективной рекламы важно чтобы максимально большой процент аудитории блогера был релевантным к рекламируемому товару. Чем уже специализация блогера и выше процент вовлеченности его аудитории, тем сильнее данный лидер мнения влияет на свою аудиторию.

## 2 АНАЛИЗ ДАННЫХ

### 2.1 Предобработка

Набор данных, с которым необходимо работать, представляет собой совокупность значений и параметров, размерность которых составляет 7 параметров. Общая совокупность выборки разделена на 2 подвыборки: тренировочную и тестовую. Пример тренировочных данных представлен на рисунке 3.

	подписки	фото	видео	карусель	длина поста	Лайки	Комментарии
count	4.690000e+02	469.000000	469.000000	469.000000	469.000000	469.000000	469.000000
mean	1.008602e+06	0.473348	0.371002	0.155650	1234.797441	51873.326226	1087.618337
std	7.300798e+05	0.499822	0.483589	0.362911	746.387651	48956.893712	2092.129556
min	3.370000e+02	0.000000	0.000000	0.000000	2.000000	106.000000	1.000000
25%	7.903060e+05	0.000000	0.000000	0.000000	460.000000	26930.000000	365.000000
50%	8.706890e+05	0.000000	0.000000	0.000000	1370.000000	36488.000000	671.000000
75%	9.887600e+05	1.000000	1.000000	0.000000	1972.000000	60360.000000	1192.000000
max	3.429474e+06	1.000000	1.000000	1.000000	2199.000000	327819.000000	33923.000000

Рисунок 3 – Общий вид данных

В результате, общий набор данных составляет две совокупностями: тренировочный (размерность 469x7) и тестовый (размерность 121x7).

Из предварительного анализа можно заметить, что в тренировочном наборе количество наблюдений больше, чем количество параметров, описывающих их. Это нормально

Следующим шагом стала проверка каждого поля каждого наблюдения на наличие параметров с отсутствующими значениями более чем 1%. и для работы с ними был применен метод средних значений.

Также данные были разделены на несколько (общая статистика блогера, количество постов и т.д.) таблиц их пришлось приводить в одну. Была проблема сопоставить общую статистику и посты так как в общей статистике были данные по лайкам и комментариям за день по всем постам, а количество постов в день могло достигать 5 штук, а в таблице посты каждый пост был отдельно и указывалось конкретное количество откликов пользователя на пост. Для работы модели требовалось указать количество подписчиков на день в который был сделан пост, а данные были только в общей статистике пришлось писать скрипт для сопоставления даты поста и количества подписчиков в этот день для наполнения таблиц.

Когда все операции по проверке данных были проведены, был построен график распределения значений. Результат представлен на рисунке 4.

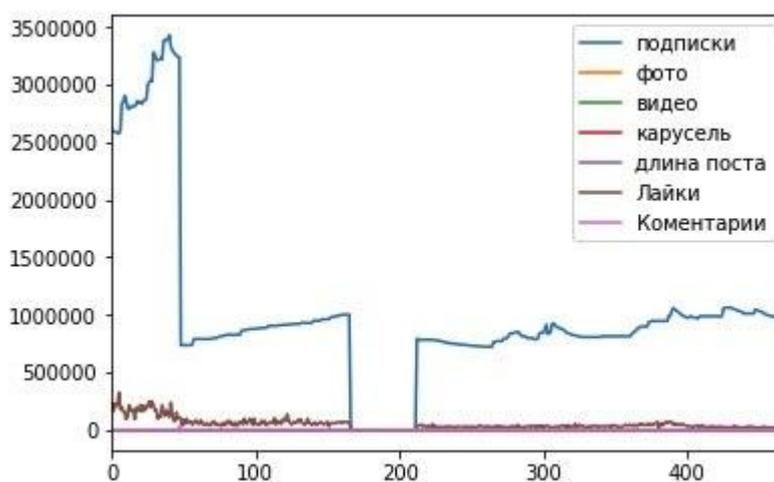


Рисунок 4 – Распределение данных

## 2.2 Оценка комплаентности аудитории

Комплаентность (в медицине) – точное и осознанное выполнение рекомендаций врача. Принято считать, что комплаентность пациента изначально максимальная и уменьшается по мере столкновения с негативными факторами (высокой стоимостью лечения, непрофессиональным поведением врача и т.д.).

В медицине комплаентность определяется множеством факторов, среди которых:

- Удобство приёма препарата.
- Озабоченность собственным заболеванием.
- Контакт врача и пациента.
- Престиж лечения и медицины в обществе в целом и в глазах пациента в частности.
- Вторичные выгоды от заболевания (например, получение денежных выплат, пособий, льгот по инвалидности) будут снижать комплаентность.
- Психические расстройства любого уровня.
- Тяжесть побочных эффектов на фоне приема препарата, побуждающих пациента прекратить терапию.

Применительно к социальным сетям – комплаентность – это склонность доверять мнению высказанному блогером и следовать его советам в реальной

жизни, принимать его как свое. Комплаентность аудитории -  $K$  - процент подписчиков, принимающих мнение блогера. Изначально в социальных сетях комплаентность аудитории блогера стремится к нулю и способна меняться в зависимости от различных факторов.

Составим упрощённую модель и обозначим комплаентность аудитории через параметры:

- Вовлеченность аудитории –  $ER$
- Тематика блога -  $T = \{T_1, T_2, \dots, T_k\}$
- $a_i$  – это вес темы освещаемая блогером
- Коэффициент автосинхронизации –  $sh = f(N)$ ,  $N = 0, 1, 2, 3, \dots$  – количество подписчиков;  $f$  - непрерывная функция, с отрицательной второй производной (HARA) – Функция полезности
- $M$  - это количество блогеров. Блогеры имеют разные тематики, но при этом кто-то может иметь несколько разных тематик одновременно. То есть если один блогер занял тематику, то эту тематику не может использовать другой.  $Z_i$  – это тематики, о которых пишет  $i$ -ый блогер.

При этом итоговая комплаентность ( $K$ ):

$$K(Blogger_i) = \frac{f(N_i) * ER_i * \sum_{j \in Z_i} I_{\{T_j\}} * a_i}{\sum_{j \in \{1, 2, \dots, M\}} (\sum_{k \in Z_j} I_{\{T_k\}} * a_k) * ER_j * f(N_j)}, \quad i=1, 2, 3, \dots, M \quad (10)$$

Формула 10. - комплаентность аудитории блогера

Вовлеченность аудитории это общепринятая маркетологами характеристика блога, вычисляющаяся по формуле  $ER = (\text{лайки} + \text{комментарии}) / \text{подписчики} * 100\%$ .

Таким образом Формула 10 отображает зависимость комплаентности аудитории от заданных параметров.

## 3 ФОРМИРОВАНИЕ МОДЕЛЕЙ И РЕЗУЛЬТАТ РАБОТЫ

### 3.1 Нейронная сеть

При выборе блогера неопытные маркетологи ориентируются на количество подписчиков у блогера, однако этого недостаточно. Одним из важных параметров является ER, но в разных социальных сетях разные способы отклика аудитории и лайков с комментариев не всегда хватает. Для этого следующим шагом с помощью нейронной сети мы будем предсказывать ER.

Основные параметры нейронной сети были описаны в главе 1. Основной задачей данного шага является нахождение оптимального набора параметров.

Список параметров, которые были определены в ходе проверки кросс валидации.

Таблица 2 – Основные параметры Нейронной сети

Параметр	Описание
Количество нейронов в скрытом слое n_hidden	Параметр влияет на сложность модели. Варьируется от 30 до 50 с шагом 2.
Количество скрытых слоев n_layers	Параметр влияет на сложность модели. Варьируется от 1 до 2.
Функции активации activation	Определяет сходимость и скорость сходимости ИНС. Функции, которые были рассмотрены: sigmoid, tanh, ReLU(LeakyReLU), ELU
Размер пакета batch_size	Размер пакета при обучении сети методом пакетного градиентного спуска. Принимает значения 18, 32, 54.

Аналогично, при выборе подходящей комбинации параметров использовался метод скользящего контроля по 5 блокам. В результате, оптимальными параметрами сети стали следующие значения:

Таблица 3 – Оптимальные параметры Нейронной сети

Параметр	Значение
n_hidden	36
n_layers	2
activation	ReLU
batch_size	32

Одним из преимуществ использования Tensorflow является поставляемая с ней утилита Tensorboard, которая позволяет визуализировать граф вычислений и процесс обучения [23]. Общий вид ИНС показан на рисунке 5.

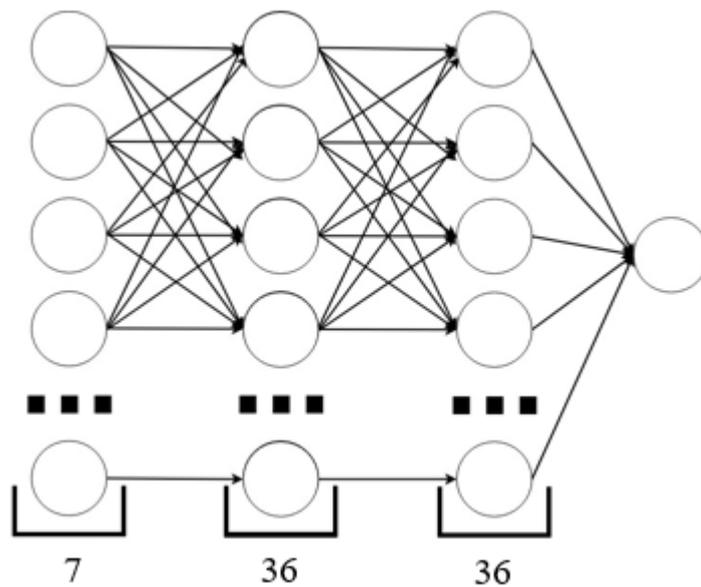


Рисунок 5 – Общий вид нейронной сети

При данных параметрах тестовая ошибка Mse составляет 35. График процесса обучения показан на рисунке 20.

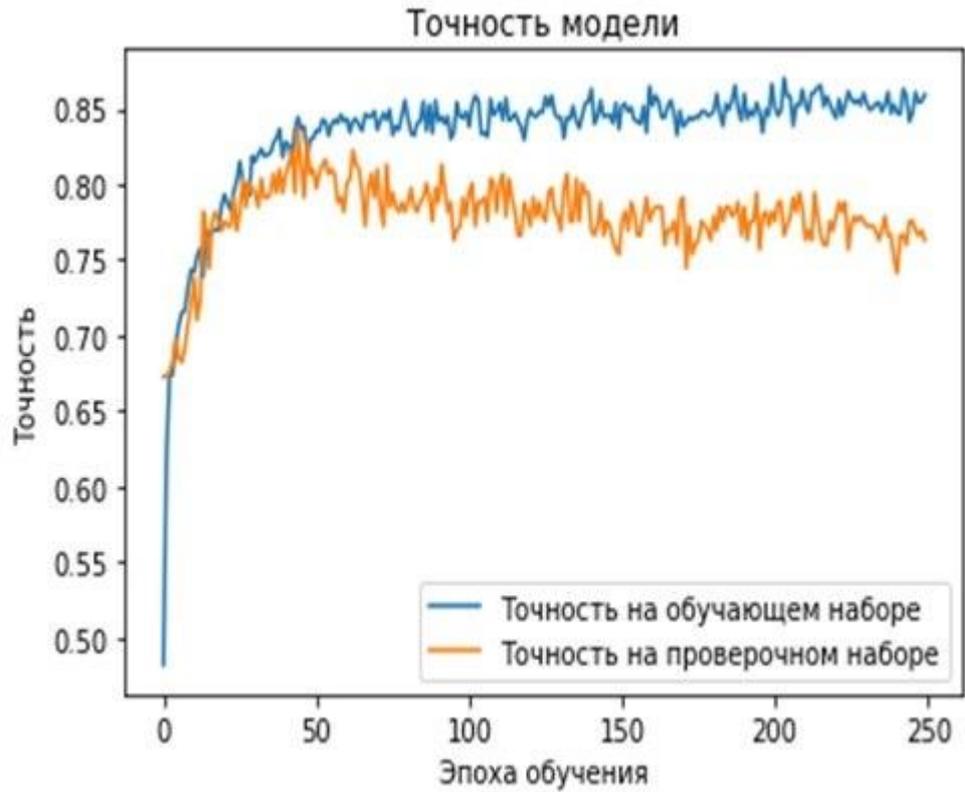


Рисунок 6 – График процесса обучения

```

469/469 [=====] - 0s 57us/sample - loss: 36.8372 - acc: 0.9041
Epoch 10/1000 |
469/469 [=====] - 0s 57us/sample - loss: 36.7861 - acc: 0.9041
Epoch 11/1000
469/469 [=====] - 0s 55us/sample - loss: 36.7321 - acc: 0.9184
Epoch 12/1000
469/469 [=====] - 0s 60us/sample - loss: 36.7436 - acc: 0.9062
Epoch 13/1000
469/469 [=====] - 0s 57us/sample - loss: 36.6497 - acc: 0.9126
Epoch 14/1000
469/469 [=====] - 0s 57us/sample - loss: 36.7049 - acc: 0.9041
Epoch 15/1000
469/469 [=====] - 0s 55us/sample - loss: 36.5554 - acc: 0.9184
Epoch 16/1000
469/469 [=====] - 0s 60us/sample - loss: 36.5211 - acc: 0.9041
Epoch 17/1000
469/469 [=====] - 0s 55us/sample - loss: 36.5012 - acc: 0.9062
Epoch 18/1000
469/469 [=====] - 0s 57us/sample - loss: 36.4655 - acc: 0.9184
Epoch 19/1000
469/469 [=====] - 0s 55us/sample - loss: 36.4498 - acc: 0.9126
Epoch 20/1000
469/469 [=====] - 0s 57us/sample - loss: 36.3998 - acc: 0.9184
Epoch 21/1000
469/469 [=====] - 0s 55us/sample - loss: 36.4005 - acc: 0.9062
Epoch 22/1000
█ █ █

Epoch 994/1000
469/469 [=====] - 0s 53us/sample - loss: 35.7028 - acc: 0.9147
Epoch 995/1000
469/469 [=====] - 0s 51us/sample - loss: 35.7235 - acc: 0.9184
Epoch 996/1000
469/469 [=====] - 0s 53us/sample - loss: 35.8685 - acc: 0.9190
Epoch 997/1000
469/469 [=====] - 0s 49us/sample - loss: 35.8546 - acc: 0.9041
Epoch 998/1000
469/469 [=====] - 0s 51us/sample - loss: 35.7177 - acc: 0.9062
Epoch 999/1000
469/469 [=====] - 0s 49us/sample - loss: 35.7003 - acc: 0.9184
Epoch 1000/1000
469/469 [=====] - 0s 51us/sample - loss: 35.6800 - acc: 0.9184

```

Рисунок 7 – Процесса обучения

Результат работы нейросети хуже, в основном по причине того, что при предобработке данных важность признаков вычислялась на основе алгоритмов бустинга. Еще одним моментом худшей производительности ИНС является малое количество данных.

Кроме оптимальных параметров нейронной сети, были опробованы и другие с изменением: количества нейронов в скрытом слое, количество скрытых слоев и функции активации.

Некоторые из них представлены на рисунках 8, 9, 10.

```
Epoch 1/1000
469/469 [=====] - 0s 285us/sample - loss: 40.3873 - acc: 0.8204
Epoch 2/1000
469/469 [=====] - 0s 68us/sample - loss: 40.2546 - acc: 0.8126
Epoch 3/1000
469/469 [=====] - 0s 62us/sample - loss: 40.2413 - acc: 0.8136
Epoch 4/1000
469/469 [=====] - 0s 57us/sample - loss: 40.2029 - acc: 0.8047
Epoch 5/1000
469/469 [=====] - 0s 55us/sample - loss: 40.2615 - acc: 0.8104
Epoch 6/1000
469/469 [=====] - 0s 55us/sample - loss: 40.8573 - acc: 0.8112
Epoch 7/1000
```

Рисунок 8 – ReLU\_2 слоя\_48 нейронов

```
469/469 [=====] - 0s 51us/sample - loss: 49.7235 - acc: 0.7104
Epoch 996/1000
469/469 [=====] - 0s 53us/sample - loss: 49.8685 - acc: 0.7190
Epoch 997/1000
469/469 [=====] - 0s 49us/sample - loss: 49.8546 - acc: 0.7041
Epoch 998/1000
469/469 [=====] - 0s 51us/sample - loss: 49.7177 - acc: 0.7062
Epoch 999/1000
469/469 [=====] - 0s 49us/sample - loss: 49.7003 - acc: 0.7104
Epoch 1000/1000
469/469 [=====] - 0s 51us/sample - loss: 49.6800 - acc: 0.7104
```

Рисунок 9 – ELU\_2 слоя\_54 нейронов

```

469/469 [=====] - 0s 51us/sample - loss: 64.6764 - acc: 0.5041
Epoch 952/1000
469/469 [=====] - 0s 55us/sample - loss: 64.6768 - acc: 0.5083
Epoch 953/1000
469/469 [=====] - 0s 51us/sample - loss: 64.6767 - acc: 0.5083
Epoch 954/1000
469/469 [=====] - 0s 51us/sample - loss: 64.6746 - acc: 0.5104
Epoch 955/1000
469/469 [=====] - 0s 51us/sample - loss: 64.6748 - acc: 0.5062
Epoch 956/1000
469/469 [=====] - 0s 53us/sample - loss: 64.6763 - acc: 0.5104
Epoch 957/1000
469/469 [=====] - 0s 51us/sample - loss: 64.6749 - acc: 0.5041
Epoch 958/1000
469/469 [=====] - 0s 51us/sample - loss: 64.6748 - acc: 0.5041
Epoch 959/1000

```

Рисунок 10 – Tanh\_3 слоя\_30 нейронов

Хуже всего себя показала функция активации tanh 3-мя слоями и 30 нейронами в каждом слое, ошибка составила в районе loss: 64.

В итоге была сделана попытка прогнозировать параметр ER для дальнейшего прогнозирования комплаентности аудитории блогера.

### 3.2 Нечёткая логика

Для прогнозирования комплаентности аудитории блогера (далее К). Было принято решение воспользоваться нечёткой логикой т.к. однозначно тяжело понять точную зависимость параметров относительно друг друга.

Листинг:

```

#Значение которые могут принимать параметры
ER = ctrl.Antecedent(np.arange(0, 101, 1), 'ER')
sh = ctrl.Antecedent(np.arange(0, 6, 1), 'sh')
K = ctrl.Consequent(np.arange(0, 101, 1), 'K')

#Правила нечёткого вывода
rule1 = ctrl.Rule(ER['poor'] & sh['poor'], K['low'])
rule2 = ctrl.Rule(ER['poor'] & sh['average'], K['low'])
rule3 = ctrl.Rule(ER['poor'] & sh['good'], K['medium'])
rule4 = ctrl.Rule(ER['average'] & sh['poor'], K['low'])
rule5 = ctrl.Rule(ER['good'] & sh['poor'], K['low'])
rule6 = ctrl.Rule(ER['average'] & sh['poor'], K['low'])
rule7 = ctrl.Rule(ER['average'] & sh['average'], K['medium'])

```

```

rule8 = ctrl.Rule(ER['average'] & sh['good'], K['high'])
rule9 = ctrl.Rule(ER['good'] & sh['average'], K['medium'])
rule10 = ctrl.Rule(sh['good'] & ER['good'], K['high'])

```

Ниже представлены результаты работы программы для двух случаев из таблицы 4. Рисунки 11 и 12.

Таблица 4 – Работа нечёткой логики

Переменные	Эксперимент №1	Эксперимент №2
ER	6.5	20
sh	2	3
К	28.08	42.54

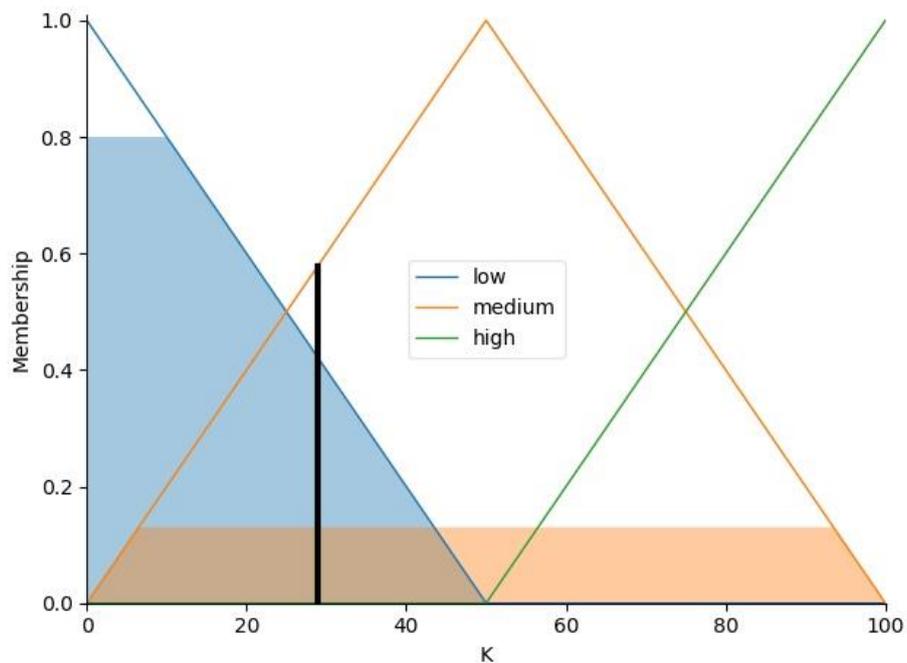


Рисунок 11 – функция принадлежности переменной К

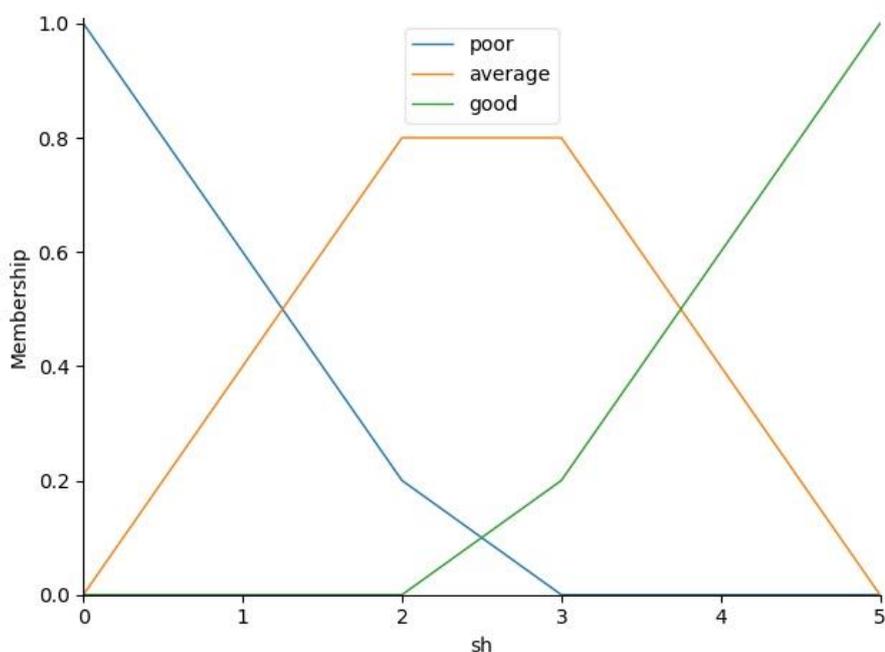


Рисунок 12 – функция принадлежности переменной Sh

Теперь мы можем сравнить комплаентность двух подопытных блогеров:

Таблица 5 – Сопоставление K и доход от продажи курсов по доп. образованию.

Переменные	@alexandramitroshina	@anarbachoo
ER	2.37	2.9
sh	4	2
K	46.3	23.3
Рублей/месяц	5 000 000	1 500 000

Мы видим, что комплаентность различается почти в 2 раза, а доход в 3.25 раз.

## **4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

### **Введение**

На сегодняшний день научные исследования характеризуются в большей степени не только объемом работ и значимостью открытия, оценку которому на начальных этапах дать достаточно трудно, а также коммерческой составляющей проекта. Специалисты высокого уровня обладают качествами как технического, так и экономического профиля, способные привести обоснования финансовой эффективности своих технологических решений. Таким образом, задачей раздела «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» является проектирование систем, отвечающих требованиям перспективности, ресурсоэффективности и экономичности разработки, а также приведение необходимых расчетов для оценки ее коммерческого потенциала.

Цель данного раздела – проектирование и создание конкурентоспособных разработок и технологий, которые будут отвечать современным требованиям в областях ресурсоэффективности и ресурсосбережения.

Данный раздел выпускной квалификационной работы включает в себя оценку коммерческого потенциала создания проекта предсказания банковских транзакций на основе методов машинного обучения. В данной части происходит определение возможных альтернатив разработки, отвечающих требованиям финансовой и ресурсной эффективности.

Конечным продуктом анализа является программная реализация созданного алгоритма. В разделе представлен анализ готового к реализации проекта, а также оценка альтернативных решений создания проекта. Также произведен расчет ресурсной и финансовой эффективности исследования.

### **4.1 Потенциальные потребители результатов исследования**

Целевой аудиторией данной системы скоринга являются:

- Рекламодатели
- Продюсеры

- Маркетинговые агентства

Целевым рынком разработки является рынок рекламы в социальных сетях. Один из критериев сегментации – вид потребителей, основанный на целевой аудитории. Второй критерий – размеры компании-заказчика.

В таблице 1 представлена карта сегментирования рынка на основе наиболее значимых критериев.

Таблица 1 – Карта сегментирования рынка прогнозирующих систем:

Критерий		Вид компании		
		Рекламодатели	Маркетинговые агентства	Продюсеры
Размер компании	Крупные			
	Средние			
	Мелкие			

Церебро      Таргет Хантер      Facebook

В результате анализа и сегментирования рынка прогнозирующих систем получается, что основным сегмента рынка, на который необходимо ориентироваться, является область мелких и средних рекламодатели и маркетинговые агентства.

#### 4.2 Анализ конкурентных технических решений

В данный момент на рынке существуют следующие аналоги разрабатываемой системы (ПО):

- Программный продукт «Церебро» (К1);
- Программный продукт «Таргет Хантер» (К2);
- Программный продукт «Facebook» (К3).

Оценочная карта для сравнения конкурентных программных решений представлена в таблице 2.

Таблица 2 – Оценочная карта для сравнения конкурентных программных решений

Критерии оценки	Вес критерия	Баллы				Конкурентоспособность			
		Бф	Бк1	Бк2	Бк3	Кф	Кк1	Кк2	Кк3
<b>Технические критерии оценки ресурсоэффективности</b>									
Наличие функционала	0,2	4	3	4	4	0,8	0,6	0,8	0,8
Устойчивость (способность обеспечивать продолжение работы после возникновения отклонений)	0,2	4	5	5	5	0,8	1	1	1
Уровень защиты доступа	0,1	5	5	5	5	0,5	0,5	0,5	0,5
Простота интерфейса	0,15	3	3	2	3	0,45	0,45	0,3	0,45
<b>Экономические критерии оценки ресурсоэффективности</b>									
Конкурентоспособность продукта	0,1	1	3	4	4	0,1	0,3	0,4	0,4
Область применения (широта функционала)	0,15	2	3	4	3	0,3	0,45	0,6	0,45
Уровень проникновения на рынок	0,05	0	3	3	4	0	0,15	0,15	0,2
Поддержка продукта	0,05	1	3	3	5	0,05	0,15	0,15	0,25
<b>Итого</b>	<b>1</b>					3	3.6	3.9	4.05

Экспертная оценка основных технических и экономических характеристик конкурентных программных решений показывает, что разрабатываемая система скоринга блогеров является конкурентоспособной по сравнению с представленными аналогами.

Основными достоинствами конкурентных программных продуктов являются большой функционал и области применения, а недостатками достаточно сложный пользовательский интерфейс, требующий некоторого обучения.

Разрабатываемая система обладает более специализированным функционалом и имеет простой пользовательский интерфейс.

### 4.3 SWOT-анализ

Для комплексного анализа научно-исследовательского проекта на основе анализа конкурентных решений была составлена матрица SWOT-анализа, содержащая сильные и слабые стороны проекта, а также возможности и угрозы для разработки проекта. Данная матрица представлена в приложении Б.

Для того, чтобы разобраться с различными комбинациями взаимосвязей областей матрицы SWOT, были построены интерактивные матрицы проекта, показывающие соответствия параметров SWOT-анализа.

Матрица SWOT-анализа

	<p><b>Сильные стороны научно-исследовательского проекта:</b></p> <p>С1. Актуальность разработки.</p> <p>С2. Простой интерфейс пользователя.</p> <p>С3. Гибкая настройка времени параметров модели.</p> <p>С4. Возможность взаимодействия системы с различными типами данных.</p> <p>С5. Высокая точность работы.</p>	<p><b>Слабые стороны научно-исследовательского проекта:</b></p> <p>Сл1. Нестабильная работа при взаимодействии с различными антивирусными программами.</p> <p>Сл2. Необходима постоянная поддержка продукта.</p> <p>Сл3. Сложное для пользователя добавление дополнительных данных для прогноза.</p> <p>Сл4. Небольшой опыт разработки.</p> <p>Сл5. Незнание программного продукта на целевом рынке.</p>
<b>Возможности:</b>	<b>Направления развития:</b>	<b>Сдерживающие факторы:</b>

<p>V1. Тенденция роста спроса на программный продукт.  V2. Рост потребностей клиентов.  V3. Захват смежных сегментов целевого рынка.</p>	<p>1. V1B2C1C3 – разработка различных версий программного продукта, ориентированных на более узкую целевую аудиторию.  2. V2B3C1C2C5 – Расширение списка анализируемых параметров.  3. V1B2B3C1C4 – реклама и продвижение продукта на смежном рынке.</p>	<p>1. V1B2Cл2Cл4 – расширение команды разработчиков.  2. V3Cл3Cл5 – Оптимизация продукта путем реализации гибкого интерфейса.</p>
<p><b>Угрозы:</b>  У1. Появление и развитие аналогичных систем.  У2. Непопулярность продукта на рынке.  У3. Сбои работы в различных средах функционирования продукта.</p>	<p><b>Угрозы развития:</b>  1. У1C3C4C5 – использование широкого функционала и удобного интерфейса.  2. У2C1C2 – реклама и продвижение на целевом рынке</p>	<p><b>Уязвимости:</b>  1. У3Cл1 – необходима дополнительная отладка для решения проблемы совместимости.  2. У1Cл4 – слияние с конкурирующим проектом</p>

#### 4.4 Оценка готовности проекта к коммерциализации

На какой бы стадии жизненного цикла не находилась научная разработка полезно оценить степень ее готовности к коммерциализации и выяснить уровень собственных знаний для ее проведения (или завершения). Для этого необходимо заполнить специальную форму:

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1.	Определен имеющийся научно-технический задел	3	5
2.	Определены перспективные направления коммерциализации научно-технического задела	5	4

3.	Определены отрасли и технологии (товары, услуги) для предложения на рынке	2	3
4.	Определена товарная форма научно-технического задела для представления на рынок	3	2
5.	Определены авторы и осуществлена охрана их прав	4	4
6.	Проведена оценка стоимости интеллектуальной собственности	4	3
7.	Проведены маркетинговые исследования рынков сбыта	5	2
8.	Разработан бизнес-план коммерциализации научной разработки	2	2
9.	Определены пути продвижения научной разработки на рынок	4	3
10.	Разработана стратегия (форма) реализации научной разработки	3	3
11.	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	5	4
12.	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	3	4
13.	Проработаны вопросы финансирования коммерциализации научной разработки	5	3
14.	Имеется команда для коммерциализации научной разработки	2	2
15.	Проработан механизм реализации научного проекта	2	4
	<b>ИТОГО БАЛЛОВ</b>	49	48

Исходя из значений показателей таблицы можно заметить, что значения каждого из оценочных параметров попадает в диапазон значений от 45 до 59, что является показателем перспективы выше среднего. Таким образом, разработка имеет хороший коммерческий потенциал.

#### **4.5 Коммерциализация результатов научно-технического исследования**

При коммерциализации научно-технических разработок разработчик преследует вполне определенную цель, которая во многом зависит от того, куда

в последующем он намерен направить (использовать, вложить) полученный коммерческий эффект.

При разработке системы скоринга блогеров наиболее подходящим способом коммерциализации является инжиниринг, и, в частности, эксплуатационный инжиниринг. Сфера деятельности, где необходимы знания о маркетинге в социальных сетях являются маркетинговые структуры и аналитические компании, которым необходимо иметь представление о рынке.

Предоставляя разработанную технологию и изменяя ее под конкретную область исследований, можно добиться коммерциализации продукта на постоянной основе.

## 4.6 Инициация проекта

Устав проекта документирует бизнес-потребности, текущее понимание потребностей заказчика проекта, а также новый продукт, услугу или результат, который планируется создать.

### 4.6.1 Определение возможных альтернатив научных исследований

Для определения возможных альтернативных путей проведения научных исследований использовался морфологический подход.

Морфологическими характеристиками являются:

- Вид функции активации;
- язык программирования;
- формат хранения данных;
- метод оптимизации;
- количество скрытых слоёв сети.

В таблице 8 представлена морфологическая матрица проекта.

Таблица 8 – Морфологическая матрица проекта

	Исполнение 1	Исполнение 2	Исполнение 3
А. Язык программирования	C++	Python	C#
Б. Формат изображений	CSV	PostgreSQL	MongoDB
В. Количество скрытых слоёв	3	4	5

Г. Вид функции активации	Логистическая	Пороговая	ReLU
Д. Метод оптимизации	Adam	AdaGrad	MomentumSGD

Из данной морфологической матрицы проекта было выделено три варианта решения технической задачи:

И1. А1Б3В3Г2Д1;

И2. А2Б2В3Г1Д3;

И3. А3Б3В2Г3Д1.

Эти варианты исполнения будут использованы в дальнейших расчетах.

#### 4.6.1 Цели и результат проекта

Здесь приводится информация о заинтересованных сторонах проекта, иерархии целей проекта и критериях достижения целей.

Под заинтересованными сторонами проекта понимаются лица или организации, которые активно участвуют в проекте или интересы которых могут быть затронуты как положительно, так и отрицательно в ходе исполнения или в результате завершения проекта.

<b>Заинтересованные стороны проекта</b>	<b>Ожидания заинтересованных сторон</b>
<b>Рекламодатели</b>	<b>Увеличение эффективности и охвата рекламы с помощью анализа комплоентности аудитории блогера (блога)</b>
<b>Маркетинговые агентства</b>	<b>Успешное апробирование алгоритмов скоринга. Данный алгоритм способен адекватно с определенной точностью определить эффективность воздействия блогера.</b>

Далее необходимо представить информацию о иерархии целей проекта и критериях достижения целей.

<b>Цели проекта:</b>	Разработка алгоритмов анализа популярности блогера у его целевой аудитории в социальных сетях
<b>Ожидаемые результаты проекта:</b>	Программа, способная количественно оценить популярность блогера у его аудитории.
<b>Критерии приемки результата проекта:</b>	Программа удовлетворяет критериям точности, надежности, быстродействия.
<b>Требования к результату проекта:</b>	<b>Требование:</b>
	Результат работы программы не превышает определенную экспертами или условиям самой задачи значения погрешность.
	Время работы программы не превышает определенную экспертами или условиям самой задачи время.
	Работа программы не должна прерываться при определенных параметрах системы (неточный набор входных данных и пр.)

#### 4.6.2 Организационная структура проекта

На данном этапе работы необходимо решить следующие вопросы: кто будет входить в рабочую группу данного проекта, определить роль каждого участника в данном проекте, а также прописать функции, выполняемые каждым из участников и их трудозатраты в проекте.

<b>№ п/п</b>	<b>ФИО, основное место работы, должность</b>	<b>Роль в проекте</b>	<b>Функции</b>	<b>Трудо-затраты, час.</b>
1	Погожев А. О., Магистрант	Разработчик	Разработка программной системы	600
2	Губин Е.И., Заведующий лабораторией кандидат физико-математических наук	Руководитель	Координация деятельности, проверка реализации	81
<b>ИТОГО:</b>				681

#### 4.6.3 Ограничения и допущения проекта

Ограничения проекта – это все факторы, которые могут послужить ограничением степени свободы участников команды проекта, а так же «границы

проекта» - параметры проекта или его продукта, которые не будут реализованных в рамках данного проекта.

– Фактор	– Ограничения/ допущения
– 3.1. Бюджет проекта	– <b>184 539,86</b>
– 3.1.1. Источник финансирования	– <b>НИ ТПУ</b>
– 3.2. Сроки проекта:	– <b>4 месяца</b>
– 3.2.1. Дата утверждения плана управления проектом	– <b>02.02.2019</b>
– 3.2.2. Дата завершения проекта	– <b>02.06.2019</b>
– 3.3. Прочие ограничения и допущения*	– <b>Отсутствуют</b>

## 4.7 Планирование научно-исследовательских работ

### 4.7.1 Структура работ в рамках научного исследования

Для планирования комплекса предполагаемых работ был составлен перечень этапов и работ в рамках проведения научного исследования, определены исполнители проекта и распределено время и этапы работ между исполнителями проекта. Исполнителями проекта являются студент и научный руководитель.

Код работы	Название	Дни	Дата начала работ	Дата окончания работ	Состав участников
1	Составление ТЗ	2	02.02.2019	04.02.2019	Магистрант, Руководитель
2	Поиск и исследование литературы	10	05.02.2019	15.02.2019	Магистрант
3	Выбор направления исследования	10	16.02.2019	27.02.2019	Магистрант, Руководитель
4	Теоретическое исследование и расчёты	15	28.02.2019	18.03.2019	Магистрант, Руководитель
5	Разработка алгоритмов предсказания данных	22	19.03.2019	12.04.2019	Магистрант

6	Тестирование разработанных алгоритмов	14	13.04.2019	29.04.2019	Магистрант
7	Оформление отчета по НИР	27	30.04.2019	01.06.2019	Магистрант, Руководитель
Итого:		100	02.02.2019	01.06.2019	

### Календарный план-график проведения работ

№	Вид работ	Исполнители	$T_{ki}$	Продолжительность выполнения работ									
				февр.				март	апрель		июнь		
				2	5	15	27	19	12	29	1		
1	Составление ТЗ	Руководитель, студент	2	2									
2	Поиск и исследование литературы	Магистрант (дипломник)	10		5								
3	Выбор направления исследования	Руководитель, Магистрант	10			15							
4	Теоретические исследования и расчёты	Руководитель, Магистрант	15				27						
5	Разработка алгоритмов предсказания данных	Магистрант (дипломник)	22					19					
6	Тестирование разработанных алгоритмов	Магистрант (дипломник)	14						12				
7	Оформление отчета по НИР	Руководитель, Магистрант	27							29		1	

Руководитель - ; Магистрант - ;

Перечень этапов, работ и распределение исполнителей по данным видам работ в рамках проводимого научно-исследовательского проекта представлен в приложении В на диаграмме Ганта.

#### 4.7.2 Бюджет научного исследования

При планировании бюджета научного исследования должно быть обеспечено полное и достоверное отражение всех видов планируемых расходов, необходимых для его выполнения. В процессе формирования бюджета, планируемые затраты группируются по статьям, представленным в таблице. В

эту статью включаются затраты на приобретение всех видов материалов, комплектующих изделий и полуфабрикатов, необходимых для выполнения работ по данной теме. Количество потребных материальных ценностей определяется по нормам расхода.

Таблица 9

*Группировка затрат по статьям*

Вид работ	Статьи									
	Сырье, материалы (за вычетом возвратных отходов), покупные изделия и полуфабрикаты	Специальное оборудование для научных (экспериментальных) работ	Основная заработная плата	Дополнительная заработная плата	Отчисления на социальные нужды	Научные и производственные командировки	Оплата работ, выполняемых сторонними организациями и предприятиями	Прочие прямые расходы	Накладные расходы	Итого плановая себестоимость
1.	566.5	29900	92693,55	9 269, 36	25119,95	0	3200	174	30588,87	191512,23

**4.7.2.1 Сырье, материалы, покупные изделия и полуфабрикаты**

Расчет стоимости материальных затрат производится по действующим прейскурантам или договорным ценам. В стоимость материальных затрат включают транспортно-заготовительные расходы (3 – 5 % от цены). В эту же статью включаются затраты на оформление документации (канцелярские принадлежности, тиражирование материалов).

Наименование	Марка, размер	Кол-во	Цена за единицу, руб.	Сумма, руб.
Бумага	Svetocopy	200	2	400
Ручки	Pilot	3	50	150
Всего за материалы				550
Транспортно-заготовительные расходы (3-5%)				16, 5
Итого по статье $C_m$				566.5

#### 4.7.2.2 Специальное оборудование для научных

#### (экспериментальных) работ

В данную статью включают все затраты, связанные с приобретением специального оборудования (приборов, контрольно-измерительной аппаратуры, стендов, устройств и механизмов), необходимого для проведения работ по конкретной теме. Определение стоимости спецоборудования производится по действующим прейскурантам, а в ряде случаев по договорной цене.

№ п/п	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс.руб.	Общая стоимость оборудования, тыс.руб.
1.	Процессор Athlon 200GE 4x3.2 ГГц	1	7000	7000
2.	Видеокарта MSI GeForce UltraHD	1	8000	8000
3	Блок питания SVEN 350W	1	1500	1500
4	ОЗУ DDR3 SODIMM 2Gb 1600MHz	2	700	1400
5	Материнская плата GIGABYTE DESIGNARE	1	5000	5000
6	Жесткий диск ASU900SS-256GM-C	1	2000	2000
7	Монитор Acer K272HLEbid Black	1	5000	5000
Итого по статье $C_o$				29900

#### 4.7.2.3 Основная заработная плата

Статья включает основную заработную плату работников, непосредственно занятых выполнением проекта, (включая премии, доплаты) и дополнительную заработную плату.

$$C_{зп} = Z_{осн} + Z_{доп}$$

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

$Z_{доп}$  – дополнительная заработная плата.

Основная заработная плата ( $Z_{осн}$ )

$$Z_{\text{осн}} = Z_{\text{дн}} \cdot T_{\text{раб}}$$

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

$T_{\text{раб}}$  – продолжительность работ, выполняемых научно-техническим работником, раб.дн.

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} \cdot M}{F_{\text{д}}}$$

где  $Z_{\text{м}}$  – месячный должностной оклад работника, руб.;

$M$  – количество месяцев работы без отпуска в течение года: при отпуске в 24 раб. дня  $M = 11,2$  месяца, 5-дневная неделя; при отпуске в 48 раб. дней  $M = 10,4$  месяца, 6-дневная неделя;

$F_{\text{д}}$  – действительный годовой фонд рабочего времени научно-технического персонала, раб.дн.

Месячный должностной оклад работника:

$$Z_{\text{м}} = Z_{\text{б}} \cdot (k_{\text{пр}} + k_{\text{д}}) \cdot k_{\text{р}}$$

где  $Z_{\text{б}}$  – базовый оклад, руб.;

$k_{\text{пр}}$  – премиальный коэффициент;

$k_{\text{д}}$  – коэффициент доплат и надбавок;

$k_{\text{р}}$  – районный коэффициент, равный 1,3 (для Томск).

На настоящем этапе сформирована команда из ключевых специалистов во главе с руководителем, имеющим опыт реализации подобных проектов. Расчёт стоимости их услуг представлен в таблице ниже:

Таблица 10 – Баланс рабочего времени

Показатели рабочего времени	Руководитель	Магистрант
Календарное число дней	365	365
Количество нерабочих дней		
• выходные дни	52	52
• праздничные дни	14	14
Потери рабочего времени		
• отпуск	48	48
• невыходы по болезни	0	0
Действительный годовой фонд рабочего времени	251	251

Расчёт основной заработной платы исполнителей системы выбирается на основе системы оплаты труда в ТПУ (для руководителя). Для исполнителя (магистра) предусматривается расчёт оплаты труда исходя из системы оплаты труда предприятия.

Таблица 11 – Расчёт основной заработной платы

Исполнители	З <sub>б</sub> , руб.	k <sub>р</sub>	З <sub>м</sub> , руб.	З <sub>дн</sub> , руб.	T <sub>р</sub> , раб.дн.	З <sub>осн</sub> , руб.
Руководитель	33 664	1,3	43 763,2	1 813,3	13,5	24 479,55
Магистрант	12 664	1,3	16 463,2	682,14	100	68 214
Итого						92 693,55

$$З_{осн} = 92\,693,55 \text{ руб.}$$

С учётом основной заработной платы, можно посчитать дополнительную заработную плату в размере 10 % от основной:

$$З_{доп} = 92\,693,55 * 0,1 = 9\,269,36$$

#### 4.7.2.4 Отчисления на социальные нужды

При начислении зарплаты работникам ежемесячно производится оплата страховых отчислений в пенсионный фонд, медицинского и социального страхования. На сегодняшний день общая ставка для всех перечисленных отчислений в России составляет 27,1%. В таблице 12 перечислены отчисления на каждого из работников.

Таблица 12 – Отчисления на социальные нужды

Исполнитель по категориям	Зар.плата,руб.	Страх. отчисления, руб.
Руководитель	24 479,55	6 633,96
Магистрант	68 214	18 485,99
Итого:	92 693,55	25 119,95

$$С_{внеб} = 25\,119,95 \text{ руб.}$$

#### 4.7.2.5 Накладные расходы

$$С_{накл} = (З_{доп} + З_{осн}) * 30\% = (92\,693,55 + 9\,269,36) * 30\% = 30\,588,87$$

#### **4.7.2.6 Оплата работ, выполняемая сторонними организациями и предприятиями**

В ходе реализации проекта были использованы услуги Internet. Для оказания подобного рода услуг, был заключён договор со сторонней организацией.

Договором установлена ежемесячная плата за услуги пользования 4G-Интернетом, составляющая 800 руб./мес.

Общее количество рабочих дней равно 94 ( $\approx 4$  мес.).

Таким образом, затраты на использования услуг, которые оказываются сторонними организациями составляют:

$$C_{\text{контр}} = 800 \cdot 4 = 3200 \text{ руб.}$$

#### **4.7.2.7 Затраты на электроэнергию**

Затраты по данной статье включают затраты на электроэнергию компьютера в процессе разработки программного продукта.

Цена электричества составляет 5,8 руб./кВт·час.

Мощность ноутбука примерно составляет 0,05 кВт.

Рабочий день составляет 6 часов. Общее количество рабочих дней равно 100.

Исходя из данных потребления электроэнергии, затраты составят:

$$C_{\text{электр}} = 5,8 \cdot 0,05 \cdot 6 \cdot 100 = 174 \text{ руб.}$$

#### **3.2.8 Итоговые затраты**

Итоговый бюджет системы состоит из затрат на сырье, заработную плату и накладных расходов. Накладные расходы составляют 16% от предыдущих статей.

$$C_{\text{проекта}} = C_{\text{м}} + Z_{\text{осн}} + Z_{\text{доп}} + C_{\text{внеб}} + C_{\text{о}} + C_{\text{контр}} + C_{\text{электр}} + C_{\text{накл}}$$
$$C_{\text{проекта}} = 566,5 + 92\,693,55 + 9\,269,36 + 25\,119,95 + 29\,900 + 3200 + 174 + 30\,588,87$$

$$C_{\text{проекта}} = 191\,512,23 \text{ руб.}$$

### 4.7.3 Реестр рисков проекта

Идентифицированные риски проекта включают в себя возможные неопределенные события, которые могут возникнуть в проекте и вызвать последствия, которые повлекут за собой нежелательные эффекты.

№	Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска*	Способы смягчения риска	Условия наступления
1	Некорректная оценка	Отказ от проекта	3	5	Высокий	Получить больший набор данных	Наличие нестандартных потоков данных
2	Закрытие предприятия клиента	Отказ от проекта	1	5	Средний	Диверсификация клиентской базы	Банкротство предприятия клиента
3	Создание конкурентами более продуманного проекта	Замещение продукта аналогом	3	4	Высокий	Постоянная доработка проекта, исследования и реализация новых подходов	Создание аналогового продукта другими разработчиками

### 4.7.4 Определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования

Разница затрат по каждому исполнению НТИ представлена в таблице 13.

Табл. 13. Разница затрат по каждому исполнению НТИ

Наименование статьи	Сумма руб.		
	Исп. 1	Исп. 2	Исп. 3
Затраты на спец. оборудование	29900	30583	41098

Интегральный финансовый показатель рассчитывается как:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{p,i}}{\Phi_{\text{max}}} \quad (13)$$

Интегральный показатель ресурсоэффективности можно определить следующим образом:

$$I_{p,i} = \sum a_i b_i \quad (14)$$

где  $I_{p,i}$  – интегральный показатель ресурсоэффективности для  $i$ -го варианта разработки,

$a_i$  – весовой коэффициент  $i$ -го варианта разработки,

$b_i$  – бальная оценка  $i$ -го варианта исполнения разработки, устанавливаемая экспертным путем по выбранной шкале оценивания,

$n$  – число параметров сравнения.

Расчет интегральных показателей ресурсоэффективности приведен в таб.

14:

Табл. 14. Расчет интегральных показателей ресурсоэффективности

Критерии	Весовой коэф.	Исп.1	Исп.2	Исп.3
Скорость работы	0,3	4	5	2
Гибкость архитектуры	0,25	5	5	5
Удобство эксплуатации	0,1	5	5	3
Потребность в ресурсах	0,1	3	4	3
Функциональные возможности	0,15	4	5	4
Итого:	1	4,25	4,9	3,25

Сравнительная эффективность разработок приведена в табл. 15:

Табл. 15. Сравнительная эффективность разработок

Показатели	Исп1	Исп2	Исп3
Интегральный финансовый показатель разработки $I_{финр}$	0,748	0,988	1
Интегральный показатель ресурсоэффективности разработки $I_p$	4,25	4,9	3,25
Интегральный показатель эффективности $I$	4,25	4,959	4,3449
Сравнительная эффективность вариантов исполнения	1	1,1668	1,02233

Исходя из проведенного анализа, можно отметить, что Исполнение №2 является более предпочтительным, нежели Исполнения №1 и №3. Несмотря на среднюю стоимость, исполнение №2 имеет наибольший показатель ресурсоэффективности. Таким образом, Исполнение №2, реализованное в данной работе, является несколько более дорогим, но и более качественным вариантом реализации проекта.

## **5 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

Так как выполнение работы заключалось в разработке системы для предсказания цифровых данных, то в качестве рабочего места будет рассмотрено рабочее место оператора персональной электронной вычислительной машины (ПЭВМ).

Использование средств вычислительной техники, накладывает целый ряд вредных факторов на человека, что впоследствии снижает производительность его труда и может привести к существенным проблемам со здоровьем сотрудника.

Обеспечение производственной и экологической безопасности является необходимым условием реализации любых проектов, в том числе конструкторских и исследовательских. В общем, обеспечение безопасности предполагает создание безопасных и благоприятных рабочих условий для лиц, задействованных в работе над проектом, а также условий, обеспечивающих экологическую безопасность окружающей среды.

Поэтому данный раздел посвящен анализу вредных и опасных факторов производственной среды для операторов ПЭВМ, разработке программ по минимизации воздействия вредоносного и опасного влияния выявленных факторов, а также программ по снижению вредных воздействий на окружающую среду, экономии невозобновимых ресурсов и защите в чрезвычайных ситуациях.

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

Регулирование отношений между работником и работодателем, касающихся оплаты труда, трудового распорядка, особенности регулирования труда женщин, детей, людей с ограниченными способностями и проч., осуществляется законодательством РФ, а именно трудовым кодексом РФ.

Режим труда и отдыха предусматривает соблюдение определенной длительности непрерывной работы на персональном компьютере (ПК) и

перерывов, регламентированных с учетом продолжительности рабочей смены, видов и категории трудовой деятельности.

Вид трудовой деятельности на персональном компьютере в рамках данной работы соответствует группе В – творческая работа в режиме диалога с ПК, категория трудовой деятельности – I (до 2 часов непосредственной работы на ПК).

При 8-часовой рабочей смене и работе на ПК, соответствующей описанным выше критериям необходимо через 1,5- 2,0 часа от начала рабочей смены и через 1,5-2,0 часа после обеденного перерыва устраивать регламентированные перерывы продолжительностью 20 минут каждый или продолжительностью 15 минут через каждый час работы.

Продолжительность непрерывной работы на ПК без регламентированного перерыва не должна превышать 2 часа.

Эффективными являются нерегламентированные перерывы (микропаузы) длительностью 1-3 минуты. Регламентированные перерывы и микропаузы целесообразно использовать для выполнения комплекса упражнений и гимнастики для глаз, пальцев рук, а также массажа. Комплексы упражнений целесообразно менять через 2-3 недели.

Продолжительность рабочего дня не должна быть меньше указанного времени в договоре, но не больше 40 часов в неделю. Для работников до 16 лет – не более 24 часов в неделю, от 16 до 18 лет и инвалидов I и II группы – не более 35 часов.

### **5.1.1 Организационные мероприятия при компоновке рабочей зоны**

Большое значение для профилактики статических физических перегрузок имеет правильная организация рабочего места человека, работающего с ПЭВМ. Рабочее место должно быть организовано в соответствии с требованиями стандартов, технических условий и (или) методических указаний по безопасности труда. Оно должно удовлетворять следующим требованиям:

обеспечивать возможность удобного выполнения работ;

учитывать физическую тяжесть работ;

учитывать размеры рабочей зоны и необходимость передвижения в ней работающего;

учитывать технологические особенности процесса выполнения работ.

Невыполнение требований к расположению и компоновке рабочего места может привести к получению работником производственной травмы или развития у него профессионального заболевания. Рабочее место программиста должно соответствовать требованиям СанПин 2.2.2/2.4.1340-03.

Конструкция оборудования и рабочего места при выполнении работ в положении сидя должна обеспечивать оптимальное положение работающего, которое достигается регулированием высоты рабочей поверхности, высоты сидения, оборудованием пространства для размещения ног и высотой подставки для ног.

На рисунке 22. схематично представлены требования к рабочему месту.

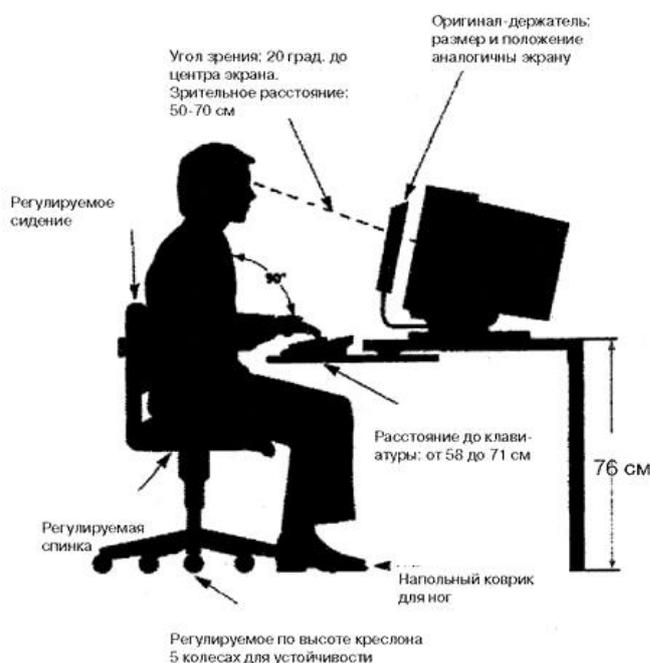


Рис. 22. Организация рабочего места

Схемы размещения рабочих мест с персональными компьютерами должны учитывать расстояния между рабочими столами с мониторами: расстояние между боковыми поверхностями мониторов не менее 1,2 м, а расстояние между экраном монитора и тыльной частью другого монитора не менее 2,0 м.

Клавиатура должна располагаться на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю. Быстрое и точное считывание информации обеспечивается при расположении плоскости экрана ниже уровня глаз пользователя, предпочтительно перпендикулярно к нормальной линии взгляда (нормальная линия взгляда 15 градусов вниз от горизонтали). Рабочие места с ПЭВМ при выполнении творческой работы, требующей значительного умственного напряжения или высокой концентрации внимания, рекомендуется изолировать друг от друга перегородками высотой 1,5 - 2,0 м.

### **5.1.2 Специфика влияния продукта на рабочий процесс**

Разрабатываемый в ходе выполнения ВКР программный продукт используется для регрессионного анализа данных. При работе с данным продуктом не требуется постоянное его использование, так как программа анализирует входные данные и получает необходимый результат, который, непосредственно, будет использован в дальнейшем. Работнику необходимо контролировать параметры входных данных и фиксировать выходные параметры. Следовательно, преимуществом данного продукта является практически полная автономность.

### **5.2 Производственная безопасность**

Выпускная квалификационная работа студента выполнялась в десятом корпусе ТПУ на кафедре информационных систем и технологий. Рабочее место находится на четвертом этаже здания и представляет собой комнату длиной – 5 м., шириной – 4 м. и высотой – 3 м. Естественное освещение кабинета осуществляется посредством одного окна размерами 2,2 м. х 1,5 м. Дверь – металлическая, одностворчатая, черного цвета. Высота двери – 2 м., ширина - 1 м. Стены комнаты окрашены водоэмульсионной краской бежевого цвета. Потолок подвесной, плиточный. Пол покрыт линолеумом. Площадь кабинета составляет 20 м<sup>2</sup>, объем – 60 м<sup>3</sup>.

Согласно СанПиН 2.2.2/2.4.1340-03, норма площади рабочего места с персональным компьютером составляет 4,5 м<sup>2</sup>. В рассматриваемой аудитории

установлено 4 рабочих места с персональными компьютерами и жидкокристаллическими экранами. Соответственно, на одного человека приходится 5 м<sup>2</sup>, что соответствует вышеуказанным требованиям. [2]

Для обеспечения производственной безопасности необходимо проанализировать воздействия на человека вредных и опасных производственных факторов, которые могут возникать при разработке или эксплуатации проекта.

Производственные условия на рабочем месте характеризуются наличием различных опасных и вредных производственных факторов, оказывающих негативное влияние на работников. В таблице 1 представлены возможные вредные и опасные факторы, возникающие при работе за ПЭВМ.

Таблица 1 – 2 Вредные и опасные факторы, возникающие при работе с ПЭВМ

Факторы (ГОСТ 12.0.003-2015)	Этапы работ			Нормативные документы
	Разрабо тка	Изготов ление	Эксплуа тация	
1. Повышенная или пониженная температура воздуха рабочей зоны	+	+	+	СанПиН 2.2.4.548–96, СанПиН 2.2.4.3359–16
2. Повышенная или пониженная влажность воздуха	+	+	+	СанПиН 2.2.2/2.4.1340–03,
3. Недостаточная освещенность рабочей зоны	+	+	+	СанПиН 2.2.1/2.1.1.1278–03, СП 52.13330.2016
4. Повышенный уровень шума на рабочем месте	+	+	+	ГОСТ 12.1.029-80 ССБТ, СН 2.2.4/2.1.8.562–96, СП 51.13330.2011
5. Повышенный уровень электромагнитных излучений	+	+	+	СанПиН 2.2.2/2.4.1340–03

6. Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	+	+	+	ГОСТ 12.1.045-84 ССБТ, ГОСТ 12.1.019-2017 ССБТ,
7. Возгорание эксплуатируемого оборудования	+	+	+	ГОСТ Р 22.0.01-2016, ГОСТ 12.1.038-82

## 5.2.1 Вредные производственные факторы

Производственный фактор считается вредным, если воздействие этого фактора на работника может привести к его заболеванию. Рассмотрим основные вредные факторы, влияющие на человека во время работы с ПЭВМ. [1]

### 5.2.1.1 Освещенность

Важнейшим фактором при создании нормальной рабочей зоны является освещение. Источники освещения могут быть искусственными или естественными. На месте рабочего места оператора задействовано смешанное освещение.

Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 - 500 лк (СНиП 23-05-95, СанПиН 2.2.2/2.4.1340-03). Освещенность поверхности экрана не должна быть более 300 лк.

Согласно СНИП 23-05-95 нормы на освещение для оператора поста управления берутся для производственных помещений. Эти нормы представлены в таблице 33.

Таблица 33 - Нормы на освещение для оператора

			<b>Искусственное освещение</b>	
--	--	--	--------------------------------	--

Характер зрительной работы	Разряд зрительной работы	Подразряд зрительной работы	Освещенность при системе общего освещения, лк	Коэффициент пульсации Кп, %	Естественное освещение КЕО $e_n$ , % при боковом
Различение объектов высокой точности	Б	1	300	15	1,0

Расчет системы искусственного освещения проводится для прямоугольного помещения, размерами: длина  $A = 5$  (м), ширина  $B = 4$  (м), высота  $H = 3$  (м).

Согласно отраслевым нормам освещенности уровень рабочей поверхности над полом составляет 0,8 (м) и установлена минимальная норма освещенности  $E = 300$  (Лк).

Световой поток ламп определяется по формуле:

$$\Phi = E_n \cdot S \cdot K_3 \cdot Z \cdot 100 / (n \cdot \eta), \quad (17)$$

Где:  $E_n$  – нормируемая минимальная освещённость по СНиП 23-05-95, (Лк);

$S$  – площадь освещаемого помещения, ( $m^2$ );

$K_3$  – коэффициент запаса, учитывающий загрязнение светильника

$Z$  – коэффициент неравномерности освещения. Для люминесцентных ламп при расчётах берётся равным  $Z = 1,1$ ;

$n$  – число светильников;

$\eta$  - коэффициент использования светового потока, (%);

$\Phi$ – световой поток, излучаемый светильником.

Подставив все значения в формулу (17), по которой рассчитывается световой поток одного источника света, получаем:

$$\Phi = \frac{300 \cdot 20 \cdot 1,5 \cdot 1,1}{4 \cdot 0,36} = 6875 \text{ (ЛМ)}$$

Выразим  $E$  из формулы (17):

$$E = \frac{(F \cdot N \cdot \eta)}{(k)} = \frac{(6875 \cdot 4 \cdot 0,36)}{(1,5 \cdot 20 \cdot 1,1)} = \frac{9900}{33} = 300 \text{ (ЛМ)} \quad (18)$$

Результаты расчёта показывают, что минимальная освещенность находится в границах нормы.

Теперь необходимо рассчитать индекс помещения:

$$i = \frac{S}{h \cdot (A + B)} \quad (19)$$

Выполним вычисления параметров с использованием следующих формул:

$$h = H - 0,8 = 3 - 0,8 = 2,2 \text{ (м)} \quad (20)$$

где  $h$  – расчетная высота подвеса светильников над рабочей поверхностью.

Экономичность осветительной установки выражается следующей формулой:

$$l = \frac{L}{h}, \quad (21)$$

где  $L$  – расстояние между рядами светильников, м.

Согласно рекомендации лампы, следует размещать параллельными рядами, принимая  $l = 1,4$ , тем самым выражение для расстояния между рядами светильников примет вид:

$$L = l * h = 1,4 * 2,2 = 3,08 \text{ (м)} \quad (22)$$

Исходя из размеров помещения, два ряда светильников будут расположены вдоль длинной стены помещения. Расстояние между двумя рядами светильников и стенами вычисляется по формуле:

$$Л = \frac{(B-L)}{2} = \frac{(4-3.08)}{2} = 0,46 \text{ (м)} \quad (23)$$

Определим индекс помещения вычисляя по формуле (19) получаем:

$$i = \frac{20}{(5 + 4) \cdot 2,2} = 1,01$$

Далее определим коэффициенты отражения поверхностей пола, потолка и стен.

Так как стены окрашены в серый цвет, являются свежепобеленными, окна без штор, следовательно, коэффициент отражения поверхности стен  $P_{ст} = 50\%$ , а коэффициент отражения поверхности потолка  $P_{п} = 30\%$  при светлой окраске потолка.

С учетом вышеперечисленных факторов и индекса помещения  $i$ , определяем значение коэффициента  $\eta = 36\%$ .

Для того, чтобы выбрать наиболее рациональные лампы, рассчитаем необходимое количество светильников по формуле:

$$N = \frac{(E \cdot k \cdot S \cdot Z)}{(n \cdot \eta \cdot F)}, \quad (24)$$

где  $E$  – норма освещенности  $E = 300$  (Лк);

$k$  – коэффициент запаса, учитывающий старение ламп и загрязнение светильников,  $k = 1,5$ ;

$S$  – площадь помещения;

$Z$  – коэффициент неравномерности освещения,  $Z = 1,1$ ;

$n$  – число рядов светильников,  $n = 2$ ;

$\eta$  – коэффициент использования светового потока,  $\eta = 0,36$ ;

$F$  – световой поток, излучаемый светильником.

Подставим численные значения в формулу (24), получим количество светильников в одном ряду:

$$N = \frac{(300 \cdot 1,5 \cdot 20 \cdot 1,1)}{(2 \cdot 6875 \cdot 0,36)} = \frac{9900}{4518} \approx 2,19 \approx 2 \text{ (шт)}$$

Таким образом, необходимо использовать в общей сложности четыре светильника со световым потоком не менее  $\Phi = 6875$  (лм). Согласно подсчитанным характеристикам наиболее оптимальной является светильник LE-СПО-03-080-1827-20X 80Вт со световым потоком 9000 (лм).

Длина одной лампы равна 1,2 (м). В светильнике находится одна лампа. При количестве светильников, равном 4, по два в двух рядах, нормы безопасности по искусственному освещению соблюдены.

Согласно расчетам, план освещения будет выглядеть как показано на рисунке 20.

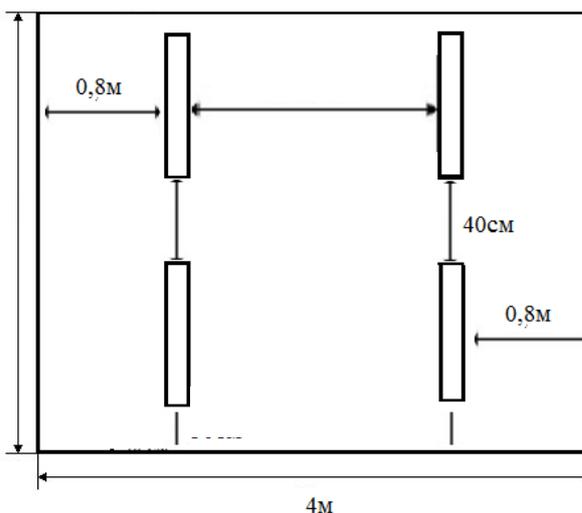


Рисунок 20 – План размещения светильников

### 5.2.1.2 Отклонение показателей микроклимата в помещении

Одним из необходимых благоприятных условий труда является обеспечение в помещениях нормальных условий микроклимата, оказывающих существенное влияние на тепловое самочувствие человека. Микроклимат в производственных помещениях, зависит от особенностей технологического процесса, а также внешних условий (категории работ, периода года, условий вентиляции и отопления).

К параметрам, характеризующим микроклимат в производственных помещениях, относятся:

- Температура воздуха ( $t$ , °C);
- Температура поверхностей ( $t$ , °C);
- Относительная влажность воздуха ( $\phi$ , %);
- Скорость движения воздуха ( $v$ , м/с);

– Интенсивность теплового облучения ( $I$ , Вт/м<sup>2</sup>).

В производственных помещениях для работы с ПЭВМ происходит постоянное выделение тепла самой вычислительной техникой, вспомогательными приборами и средствами освещения. Поскольку оператор расположен в непосредственной близости с источниками выделения тепла, то данный фактор является одним из важнейших вредных факторов производственной среды оператора ПЭВМ, а высокая температура воздуха способствует быстрому перегреву организма и быстрой утомляемости. [6]

Влажность оказывает большое влияние на терморегуляцию организма. Так, например, высокие показатели относительной влажности (более 85 %) затрудняют терморегуляцию снижая возможность испарения пота, низкие показатели влажности (менее 20 %) вызывают пересыхание слизистых оболочек человека. [7]

Работа программиста относится к категории Ia, которые производятся сидя и сопровождаются незначительным физическим напряжением. Интенсивность энергозатрат организма для данной категории работ составляет до 120 ккал/ч (до 139 Вт).

Оптимальные значения показателей микроклимата на рабочих местах производственных помещений согласно СанПиН 2.2.4.548-96 для категории работ Ia представлены в таблице 2. [3]

Таблица 2 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений

Период года	Категория работ	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia	22 – 24	60 – 40	0,1
Теплый	Ia	21 – 23	60 – 40	0,1

Оптимальные микроклиматические условия обеспечивают общее и локальное ощущение теплового комфорта при минимальном напряжении механизмов терморегуляции, не вызывают отклонений в состоянии здоровья,

создают предпосылки для высокого уровня работоспособности и являются предпочтительными на рабочих местах.

В таблице 3 приведены допустимые величины показателей микроклимата на рабочих местах производственных помещений согласно СанПиН 2.2.4.548-96 для категории работ Ia. [3]

Таблица 3 – Допустимые величины показателей микроклимата на рабочих местах производственных помещений

Период года	Категория работ	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia	20 – 25	15 – 75	0,1
Теплый	Ia	21 – 28	15 – 75	0,1 – 0,2

Допустимые микроклиматические условия не вызывают повреждений или нарушений состояния здоровья, но могут приводить к возникновению общих и локальных ощущений теплового дискомфорта, напряжению механизмов терморегуляции, ухудшению самочувствия и понижению работоспособности.

Согласно требованиям СанПиН 2.2.2/2.4.1340–03, в кабинете поддерживается температура равная 19–20 С°, при относительной влажности в 55–58%. Для этого в помещении проводится ежедневная влажная уборка и систематическое проветривание после каждого часа работы на ПЭВМ. [2]

### 5.2.1.3 Производственный шум

Шумовой фон в помещении возникает из-за работы десяти компьютеров, а также из-за принтеров, телефонов и систем вентиляции.

Во избежание негативных последствий от производственного шума, его необходимо регулировать в соответствие с нормами, которые указаны в ГОСТ 12.1.003-83 «ССБТ. Общие требования безопасности».

В соответствии с СанПин 2.2.2/2.4.1340-03 допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ приведены в таблице 3.

Таблица 3. Допустимые значения уровней звукового давления, создаваемого ПЭВМ

Уровни звукового давления в октавных полосах со среднегеометрическими частотами									Уровни звука в дБА
31,5 Гц	63 Гц	125 Гц	250 Гц	500 Гц	1000 Гц	2000 Гц	4000 Гц	8000 Гц	
86 дБ	71 дБ	61 дБ	54 дБ	49 дБ	45 дБ	42 дБ	40 дБ	38 дБ	50

Помещения, в которых для работы используются ПК, не должны граничить с помещениями, в которых уровни шума превышают нормируемые значения.

В производственных помещениях, оборудованных ПК, при выполнении основной работы на ПК уровень шума на рабочем месте не должен превышать 50 дБА.

Допустимые уровни звука на рабочих местах нормируются по

ГОСТ 12.1.003-83. Значения допустимых уровней шума приведены в таблице 4.

Таблица 4 – Допустимые уровни шума

Объект	Общий уровень звука, дБ	Уровни звукового давления, дБ в среднегеометрических частотах октавных полос, Гц							
		63	125	250	500	1000	2000	4000	8000
Постоянное рабочее место: 1) работа до 4 ч 2) работа до 8 ч	80	95	87	82	78	75	73	71	69
	86	101	93	88	81	79	77	75	

Допустимый уровень звукового давления колеблется от 38 дБ до 86 дБ при частоте от 8000 Гц до 31,5 Гц, соответственно. Для уменьшения воздействий шума можно использовать следующие методы, согласно СНиП 23-03-2003:

Экранирование рабочих мест, то есть установка перегородок между рабочими местами;

Установка оборудования, производящего минимальный шум.

Для снижения уровня шума, производимого персональными компьютерами, рекомендуется регулярно проводить их техническое

обслуживание: чистка от пыли, замена смазывающих веществ; также применяются звукопоглощающие материалы.

#### **5.2.1.4 Электромагнитное излучение**

Электромагнитное излучение - распространяющееся в пространстве возмущение электрических и магнитных полей. Источниками электромагнитного излучения в данном исследовании являются мониторы и системный блок.

Оценка величины уровней ЭМП, проведенная по паспортным данным компьютера и монитора, показала их соответствие нормам ТСО–03 и СанПиН 2.2.2/2.4.1340–03. В табл. 5 приведены нормы уровня ЭМП, которым соответствует техника в кабинете.

Таблица 5. – Допустимые уровни ЭМП, создаваемых ПК (СанПиН 2.2.2/2.4.1340–03).

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

Для того, чтобы снизить воздействие таких видов излучения, рекомендуют применять такие мониторы, у которых уровень излучения понижен (MPR-II, ТСО-92, ТСО-99), а также установить защитные экраны и соблюдать режимы труда и отдыха.

#### **5.2.2 Опасные производственные факторы**

Производственный фактор считается опасным, если его воздействие на работника может привести к его травме. [1]

### 5.2.2.1 Опасность поражения электрическим током

Поражение электрическим током является опасным производственным фактором и, поскольку оператор ПЭВМ имеет дело с электрооборудованием, то вопросам электробезопасности на его рабочем месте должно уделяться много внимания.

Опасность поражения человека электрическим током оценивается величиной тока  $I$  (А), проходящего через его тело, или напряжением прикосновения  $U$  (В). Степень опасного воздействия на человека электрического тока зависит от рода и величины напряжения тока, частоты электрического тока, пути тока через тело человека, продолжительности его воздействия на организм человека, а также условий внешней среды.

Электрический ток, протекая через тело человека, производит термическое, механическое и световое воздействие – электролитическое разложение жидкости (в том числе и крови), судорожное сокращение мышц, разрыв тканей и поражение глаз.

Работа с ПЭВМ является опасной с точки зрения поражения током, так как практически во всех частях компьютера течет электрический ток. Поражение электрическим током при работе в ПЭВМ возможно при наличии оголенных участков на кабеле, нарушении изоляции распределительных устройств и от токоведущих частей компьютера в случае их пробоя и нарушении изоляции, при работе с ПЭВМ во влажной одежде и влажными руками.

Помещение, где расположено рабочее место оператора ПЭВМ, относится к помещениям без повышенной опасности ввиду отсутствия следующих факторов: сырость, токопроводящая пыль, токопроводящие полы, высокая температура, возможность одновременного прикосновения человека к имеющим соединение с землей металлоконструкциям зданий, технологическим аппаратам, механизмам и металлическим корпусам электрооборудования.

К мероприятиям по предотвращению возможности поражения электрическим током относятся:

При производстве монтажных работ необходимо использовать только исправный инструмент, аттестованный службой КИПиА;

С целью защиты от поражения электрическим током, возникающим между корпусом приборов и инструментом при пробое сетевого напряжения на корпус, корпуса приборов и инструментов должны быть заземлены;

При включенном сетевом напряжении работы на задней панели должны быть запрещены;

Все работы по устранению неисправностей должен производить квалифицированный персонал;

Необходимо постоянно следить за исправностью электропроводки [2].

Согласно ГОСТ 12.1.038-82 на рабочем месте программиста допускаются уровни напряжений прикосновения и токов, представленные в таблице 6. [4]

Таблица 6 – Предельно допустимые напряжения прикосновения и токи

Род тока	Напряжение прикосновения, В	Ток, мА
	не более	
Переменный, 50 Гц	2,0	0,3
Постоянный	8,0	1,0

Значения напряжения прикосновения и токов приведены при продолжительности воздействия не более 10 минут в сутки.

### **5.2.2.2 Пожаровзрывобезопасность**

Возникновение пожара является опасным производственным фактором, т.к. пожар на предприятии наносит большой материальный ущерб, а также часто сопровождается травмами и несчастными случаями.

Пожарная безопасность представляет собой единый комплекс организационных, технических, режимных и эксплуатационных мероприятий по предупреждению пожаров и взрывов.

В помещениях с ПЭВМ повышен риск возникновения пожара из-за присутствия множества факторов: наличие большого количества электронных схем, устройств электропитания, устройств кондиционирования воздуха;

возможные неисправности электрооборудования, освещения, или неправильная их эксплуатация может послужить причиной пожара.

Для устранения возможных причин возникновения пожаров необходимо проводить следующие мероприятия:

- **Организационные мероприятия:**
  - противопожарный инструктаж обслуживающего персонала;
  - обучение персонала техники безопасности;
  - разработка инструкций, плакатов, планов эвакуации.
- **Эксплуатационные мероприятия:**
  - соблюдение эксплуатационных норм оборудования;
  - выбор и использование современных автоматических средств тушения пожаров.
- **Технические мероприятия:**
  - профилактический осмотр и ремонт оборудования;
  - соблюдение противопожарных мероприятий при устройстве электропроводок, оборудования, систем отопления, вентиляции и освещения.

### **5.3 Экологическая безопасность**

#### **5.3.1 Анализ воздействия продукта на окружающую среду**

Разработанный программный продукт, не наносит вреда окружающей среде ни на стадиях его разработки, ни на стадиях эксплуатации. Однако, средства, необходимые для его разработки и эксплуатации могут наносить вред окружающей среде.

Объект, на котором производилась разработка продукта, а также объекты, на которых будет производиться его использование операторами ПЭВМ относятся к предприятиям пятого класса, размер санитарно-защитной зоны для которых равен 50 м.

Основными факторами, оказывающими негативные действия на экологию, являются факторы, связанные с производством и эксплуатацией компьютерной техники. В частности, отходы и выбросы, имеющие место на этапе производства компьютеров, а также отходы, связанные с неполной их утилизацией.

Эксплуатация компьютерной техники может сопровождаться следующими негативными факторами влияния на окружающую среду:

локальное повышение электромагнитного и радиоактивного фона;

неоправданное потребление электроэнергии (связано с использованием компьютера не на полную мощность в течение всего его время работы) и прочее.

### **5.3.2 Решения по обеспечению экологической безопасности**

При разработке любых автоматизированных систем возникает необходимость утилизировать производственные отходы, в качестве которых в данном случае выступают бумажные отходы (макулатура) и неисправные детали персональных компьютеров, плат, контроллеров.

Бумажные отходы должны передаваться в соответствующие организации для дальнейшей переработки во вторичные бумажные изделия. Неисправные комплектующие персональных компьютеров должны передаваться либо организациям, осуществляющим вывоз и уничтожение бытовых и производственных отходов, либо организациям, занимающимся переработкой отходов. Важнейшим этапам обращения с отходами является их сбор, а в дальнейшем переработка, утилизация и захоронение. [8]

## **5.4 Безопасность в чрезвычайных ситуациях**

### **5.4.1 Перечень возможных ЧС при разработке и эксплуатации проектируемого решения**

Наиболее характерной для объекта, где размещаются рабочие помещения, оборудованные ПЭВМ, чрезвычайной ситуацией является пожар.

Причинами возникновения данного вида ЧС могут являться:

возникновением короткого замыкания в электропроводке;

возгоранием устройств ПЭВМ из-за неисправности аппаратуры;

возгоранием устройств искусственного освещения;

возгоранием мебели по причине нарушения правил пожарной безопасности, а также неправильного использования дополнительных бытовых электроприборов и электроустановок.

Помещение для работы операторов ПЭВМ по системе классификации категорий помещений по взрывопожарной и пожарной опасности относится к категории Д (из 5-ти категорий А, Б, В1-В4, Г, Д), т.к. относится к помещениям с негорючими веществами и материалами в холодном состоянии. [9]

Рабочее помещение расположено в 10 корпусе ТПУ, в 408 аудитории, поэтому на рисунке 1 представлен план эвакуации для этажа.

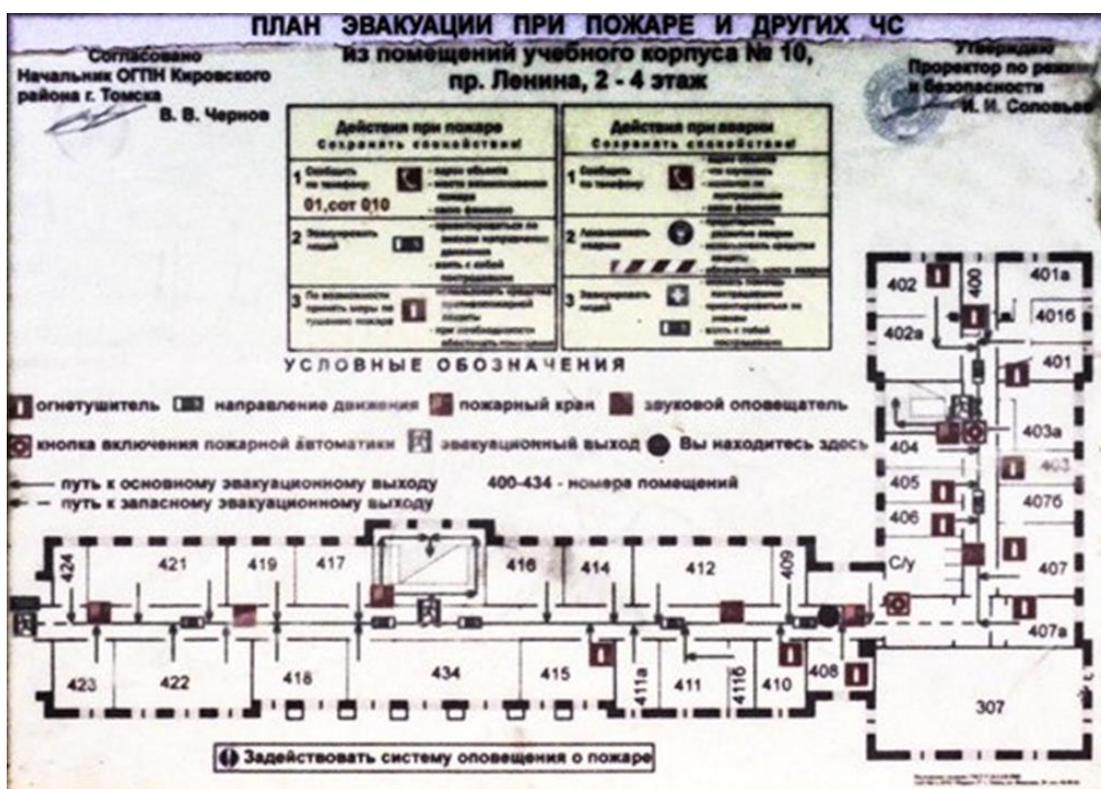


Рисунок 1. План эвакуации при ЧС 10 корпуса ТПУ.

#### 5.4.2 Разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.

Пожарная безопасность подразумевает надлежащее состояние объекта с исключением возможности возникновения очага возгорания (пожара) и его распространения в пространстве. Обеспечение пожарной безопасности — приоритетная задача для любого предприятия. Создание системы защиты регламентировано законом и нормативными документами различных ведомств.

Каждый сотрудник организации должен быть ознакомлен с инструкцией по пожарной безопасности, пройти инструктаж по технике безопасности и строго соблюдать его.

Работник при обнаружении пожара или признаков горения (задымление, запах гари, повышение температуры и т.п.) должен:

Немедленно прекратить работу и вызвать пожарную охрану по телефону «01», сообщив при этом адрес, место возникновения пожара и свою фамилию;

Принять по возможности меры по эвакуации людей и материальных ценностей;

Отключить от сети закрепленное за ним электрооборудование;

Приступить к тушению пожара имеющимися средствами пожаротушения;

Сообщить непосредственному или вышестоящему начальнику и оповестить окружающих сотрудников;

При общем сигнале опасности покинуть здание согласно «Плану эвакуации людей при пожаре и других ЧС».

## **5.5 Заключение**

Несмотря на относительную простоту проекта, рабочего пространства и степень эксплуатации, рассмотренные опасные и вредные факторы могут существенно повлиять на состояние и здоровье пользователя и окружающей среды. Были описаны основные моменты при чрезвычайных ситуациях и действия при их возникновении, а также правовые нормы и нормы эксплуатации рабочего места. Все эти замечания позволяют использовать разработанный программный комплекс эффективно и без последствий для сотрудников.

## **ЗАКЛЮЧЕНИЕ**

В результате выполнения выпускной квалификационной работы была спроектирована и реализована программа, способная для определения комплектности аудитории блогера.

Проведен анализ предметной области, технологии предобработки и формирования моделей, основные подходы к проектированию регрессионных моделей.

Были проанализированы входные данные, произведена предобработка данных путем нахождения пропущенных значений,

На основании предобработанной информации были обучена Нейронная сеть.

Для определения параметра комплаентности аудитории была обучена модель с нечёткой логикой.

Результаты работы могут быть использованы в качестве основы для дальнейшего анализа комплаентности аудитории, персонализации рекламы.

Также хочется заметить, что определение комплаентности аудитории блогера принципиально важна для эффективности рекламы и влияния блогера на аудиторию!

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Epsilon research on services [Электронный ресурс] URL: <https://us.epsilon.com/pressroom/new-epsilon-research-indicates-80-of-consumers-are-more-likely-to-make-a-purchase-when-brands-offer-personalized-experiences>  
свободный – Яз. англ. Дата обращения 02. 03.2019г.
2. Дюк Вячеслав Анатольевич, Флегонтов Александр Владимирович, Фомина Инга Константиновна Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях // Известия РГПУ им. А.И. Герцена. 2011. №138. URL: <https://cyberleninka.ru/article/n/primenenie-tehnologiy-intellektualnogo-analiza-dannyh-v-estestvennonauchnyh-tehnicheskikh-i-gumanitarnyh-oblastyah>(дата обращения:25.05.2019).
3. Aurelien Geron. Hands-On Machine Learning with Sklearn and Tensorflow. – O'Reilly, 2015: - 751p.
4. Чокоей Владимир Захарьевич Обработка и разведочный анализ числовых массивов данных // Crede Experto: транспорт, общество, образование, язык. 2017. №3. URL: <https://cyberleninka.ru/article/n/obrabotka-i-razvedochnyy-analiz-chislovyh-massivov-dannyh> (дата обращения: 21.03.2019).
5. Н. В. Решетникова, Л. А. Жуков Особенности использования предобработки данных для изучения динамики их изменения с помощью нейросетей с учителем // Новые информационные технологии в автоматизированных системах. 2005. №8. URL: <https://cyberleninka.ru/article/n/osobennosti-ispolzovaniya-predobrabotki-dannyh-dlya-izucheniya-dinamiki-ih-izmeneniya-s-pomoschyu-neyrosetey-s-uchitelem> (дата обращения:15.03.2019).
6. Bharath Ramsundar. Tensorflow for Deep Learning. – O'Reilly, 2015: - 253p.
7. Хашковский Валерий Валерьевич, Шкурко Алексей Николаевич Современные подходы в организации систем обработки больших объемов данных // Известия ЮФУ. Технические науки. 2014. №8. URL:

<https://cyberleninka.ru/article/n/sovremennye-podhody-v-organizatsii-sistem-obrabotki-bolshih-obemov-dannyh> (дата обращения: 17.03.2019).

8. Агафонов Антон Александрович, Мясников Владислав Валерьевич Оценка и прогнозирование параметров транспортных потоков с использованием композиции методов машинного обучения и моделей прогнозирования временных рядов // КО. 2014. №3. URL: [https://books.google.ru/books?hl=en&lr=&id=sm39DQAAQBAJ&oi=fnd&pg=PA11&dq=%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5+%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5&ots=lmOKvoxKbA&sig=syblyAoh40RP8zbP-dTCZM-kAmk&redir\\_esc=y#v=onepage&q=%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5%20%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5&f=false](https://books.google.ru/books?hl=en&lr=&id=sm39DQAAQBAJ&oi=fnd&pg=PA11&dq=%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5+%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5&ots=lmOKvoxKbA&sig=syblyAoh40RP8zbP-dTCZM-kAmk&redir_esc=y#v=onepage&q=%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5%20%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5&f=false) (дата обращения: 25.05.2019).

9. Базенков Николай Ильич, Губанов Дмитрий Алексеевич Обзор информационных систем анализа социальных сетей // УБС. 2013. №41. URL: <https://cyberleninka.ru/article/n/obzor-informatsionnyh-sistem-analiza-sotsialnyh-setey> (дата обращения: 01.04.2019).

10. Sebastian Rashka. Python for Machine Learning. – Packt, 2011: - 253p.

11. Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie. The elements of statistical Learning. – Packt, 2011: - 548p.

12. Christopher Bishop. Pattern Recognition and Machine Learning. – Packt, 2011: - 1148p.

13. Симонова Сабина Игоревна Интеллектуальный анализ данных для задач CRM // International Journal of Open Information Technologies. 2015. №2. URL: <https://cyberleninka.ru/article/n/intellektualnyu-analiz-dannyh-dlya-zadach-crm> (дата обращения: 04.04.2019).

14. Филатова Татьяна Владимировна Применение нейронных сетей для аппроксимации данных // Вестн. Том. гос. ун-та. 2004. №284. URL: <https://cyberleninka.ru/article/n/primenenie-neyronnyh-setey-dlya-approksimatsii-dannyh> (дата обращения: 04.04.2019).

15. Корнеев Д. С. Использование аппарата нейронных сетей для создания модели оценки и управления рисками предприятия // УБС. 2007. №17. URL: <https://cyberleninka.ru/article/n/ispolzovanie-apparata-neyronnyh-setey-dlya-sozdaniya-modeli-otsenki-i-upravleniya-riskami-predpriyatiya> (дата обращения: 04.04.2019).
16. François Chollet. Deep Learning with Python. – Packt, 2011: - 547p.
17. Архангельская Екатерина Владиславовна , Кадурич А. , Николенко Сергей Игоревич «Глубокое обучение. Погружение в мир нейронных сетей», Библиотека программиста. – 480 стр.
18. Adam Gibson and Josh Patterson. Deep Learning: A Practitioner's Approach– O’reilly, 2012: - 523p.
19. Dan Van Boxel. Hands-On Deep Learning with TensorFlow– Packt, 2014: - 174p.
20. Amita Kapoor and Antonio Gulli. TensorFlow 1.x Deep Learning Cookbook: Over 90 Unique Recipes to Solve Artificial-intelligence Driven Problems with Python – Packt, 2009: - 526p.
21. Игнатъев Н. А. О синтезе факторов в искусственных нейронных сетях // ЖВТ. 2005. №3. URL: <https://cyberleninka.ru/article/n/o-sinteze-faktorov-v-iskusstvennyh-neyronnyh-setyah> (дата обращения: 09.04.2019).
22. Antonio Gulli and Sujit Pal. Deep Learning with Keras– Packt, 2015: - 301.
23. Ahmed Menshawy, Giancarlo Zaccone, and Md. Rezaul Karim. Deep Learning with TensorFlowv– Packt, 2015: - 316p.
24. Дружков Павел Николаевич, Золотых Николай Юрьевич, Половинкин Алексей Николаевич Программная реализация алгоритма градиентного бустинга деревьев решений // Вестник ННГУ. 2011. №1. URL: <https://cyberleninka.ru/article/n/programmная-realizatsiya-algoritma-gradientnogo-bustinga-dereviev-resheniy> (дата обращения: 01.05.2019).
25. Никулин В. Н., Палешева С. А., Зубарева Д. С. Об однородных ансамблях при использовании метода бустинга в приложении к классификации несбалансированных данных // Вестник ПГУ. Серия: Экономика. 2012. №4. URL:

<https://cyberleninka.ru/article/n/ob-odnorodnyh-ansamblyah-pri-ispolzovanii-metoda-bustinga-v-prilozhenii-k-klassifikatsii-nesbalansirovannyh-dannyh> (дата обращения: 01.05.2019).

26. Охрана труда [Электронный ресурс] / Безопасность жизнедеятельности. URL: <http://www.grandars.ru/shkola/bezopasnost-zhiznedeyatelnosti/ohrana-truda.html>, свободный. Яз. Рус. Дата обращения: 16.05.2019.

27. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы. Яз. Рус. Дата обращения: 16.05.2019.

28. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений. Яз. Рус. Дата обращения: 16.05.2019.

29. ГОСТ 12.1.038–82 Система стандартов безопасности труда. Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов. Яз. Рус. Дата обращения: 16.05.2019.

30. Федеральный закон от 22.07.2008 N 123-ФЗ (ред. От 13.07.2015) «Технический регламент о требованиях пожарной безопасности» [Электронный ресурс] / КонсультантПлюс. URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_78699/](http://www.consultant.ru/document/cons_doc_LAW_78699/), свободный. Яз. Рус. Дата обращения: 16.05.2019.

31. Ефремова О. С. Требования охраны труда при работе на персональных электронно-вычислительных машинах. – 2-е изд., перераб. и доп. – М. : Издательство «Альфа-Пресс», 2008. Яз. Рус. Дата обращения: 16.05.2019.

32. Назаренко О. Б. Безопасность жизнедеятельности: учебное пособие / О. Б. Назаренко, Ю. А. Амелькович; Томский политехнический университет. – 3-е изд., перераб. и доп. – Томск: Изд-во Томского политехнического университета, 2013. Яз. Рус. Дата обращения: 16.05.2019.

33. ГОСТ Р 55090-2012 Ресурсосбережение. Обращение с отходами. Рекомендации по утилизации отходов бумаги. Яз. Рус. Дата обращения: 16.05.2019.

34. НПБ 105-03 Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности [Электронный ресурс] / Электронный фонд правовой и нормативно-технической документации. URL: <http://docs.cntd.ru/document/1200032102>, свободный. Яз. Рус. Дата обращения: 16.05.2019.

## ПРИЛОЖЕНИЕ А

(справочное)

### THEORETICAL FUNDAMENTALS

Студент:

Группа	ФИО	Подпись	Дата
8BM72	Погожев Александр Олегович		

Консультант школы отделения информационных технологий:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Мирошниченко Е. А.	к. т. н.		

Консультант – лингвист отделения информационных технологий:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ	Сидоренко Т. В.	к. п. н.		

# 1 THEORETICAL FUNDAMENTALS

## 1.1 Selection of software components

Machine learning is a rapidly developing and sought-after area of information technology. Many corporations and research centers develop algorithms and software components, as well as tools that solve common data analysis problems, based on C++ and Python.

According to market analysis, the most popular solution is to use Python + Tensorflow. When TensorFlow was open source in November 2015, there were already many popular open source libraries for deep learning and most of the functions of Tensorflow already existed in one form or another. A brief analysis of possible solutions is presented in the table below.

<b>Library</b>	<b>API</b>	<b>Platforms</b>	<b>Developer</b>	<b>Year</b>
Caffe	Python, C++, Matlab	Linux, macOS, Windows	Y. Jia, UC Berkeley (BVLC)	2013
Torch	C++, Lua	Linux, macOS, iOS, Android	R. Collobert, K. Kavukcuoglu, C. Farabet	2002
H2O	Python, R	Linux, macOS, Windows	H2O.ai	2014
MXNet	Python, C++, др.	Linux, macOS, Windows, iOS, Android	DMLC	2015
TensorFlow	Python, C++, JS, др.	Linux, macOS, Windows, iOS,	Google	2015
Theano	Python	Linux, macOS, iOS	University of Montreal	2010

However, the clean TensorFlow design, scalability, flexibility, and excellent documentation (with the support of Google) quickly raised it to the top of the list.

TensorFlow was designed to be flexible, scalable, and ready to go, and existing structures may implement only two of the three features. Some of the highlights of TensorFlow are:

- It works not only on Windows, Linux and MacOS, but also on mobile devices, including both iOS and Android.
- It provides a very simple Python API called TF.Learn (`tensorflow.contrib.learn`), compatible with Scikit-Learn. It can be used to train various types of neural networks in just a few lines of code.
- Several other high-level APIs were built independently of TensorFlow, such as Keras (now available at `tensorflow.contrib.keras`) or Pretty Tensor.
- Its core Python API offers much greater flexibility (due to higher complexity) to create all sorts of computations, including any neural network architecture that can be implemented.
- It includes highly efficient C++ implementations of many ML operations, in particular those necessary for building neural networks. There is also a C++ API for defining your own high-performance operations.
- It provides several advanced optimization nodes to search for parameters that minimize the loss function. They are very easy to use, as TensorFlow automatically takes care of calculating the gradients of the functions that the developer defines. This is called automatic differentiation.

## **1.2 Exploratory data analysis**

In statistics, exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often using visual methods. The statistical model can be used or can be rejected, but first of all the EDA is designed to see that the data can manifest, besides the formal task of modeling or testing hypotheses. John Tukey advanced research data analysis to encourage statisticians to research data and possibly formulate hypotheses that could lead to new data collection and experiments.

There are a number of tools that are useful for EDA, typical graphical methods:

- Box plot
- Histogram
- Multi-vari chart
- Scatter plot
- Stem-and-leaf plot
- Parallel coordinates
- Odds ratio
- Targeted projection pursuit
- Dimensionality reduction

### **1.2.1 Missing Parameter Values**

Missing values reflect confusion in real-world data. The reasons for their occurrence can be different - from human errors in data entry, incorrect sensor readings to software errors in the data processing pipeline.

The simplest method of processing such data is deletion. This quick solution can work in some cases, when the share of missing values is relatively low (<10%), in most cases this will lead to the loss of a large amount of useful data.

In a more general case, the standard and often very suitable approach is to replace the missing real numbers with average, median or modal values.

With the parameter values in the form of a category, the standard procedure that needs to be done is to replace the missing entry with the most common one.

### **1.2.2 Parameters with low dispersion**

Intuitively, it can be assumed that functions with small dispersion are useless and are just noise for the model . However, there are some points that need to be considered when working with similar data:

- The dispersion of an object is not unit, and by repeatedly expressing, say, the length in meters, millimeters or feet, the dispersion changes.

- The dispersion of the function does not take into account the relationship between the function and the response of the system, which is the focus of the learning models with the teacher. While the predictor may have small deviations, the relationship between the predictor and the answer can be very strong in this range.

### **1.2.3 Parameter Correlation**

The main purpose of regression analysis is to isolate the relationship between each independent variable and the dependent variable. The interpretation of the regression coefficient is that it represents the average change in the dependent variable for each change by 1 unit in the independent variable, when all other independent variables remain constant. This last part is crucial for multicollinearity.

Multicollenarity - the presence of a linear relationship between the explanatory variables (factors) of the regression model. When independent variables are correlated, this indicates that changes in one variable are associated with shifts in another variable. The stronger the correlation, the more difficult it is to change one variable without changing the other. For a model, it becomes difficult to estimate the relationship between each independent variable and the dependent variable independently, because independent variables tend to change simultaneously.

One of their effective methods is to remove highly correlated predictors from the model. Since they provide redundant information, the removal of one of the correlated factors usually does not lead to a significant decrease in  $R^2$ .

### **1.2.4 Determination of the significance of parameters**

In data mining applications, input predictors are rarely equally significant. Often, only a few of them have a significant impact on the answer; the vast majority does not matter and can be excluded. It is often useful to know the relative importance or contribution of each input variable in predicting the response.

The advantage of using ensembles of decision tree methods, such as gradient boosting, is that they can automatically provide estimates of the importance of parameters from a trained predictive model.

As a rule, importance provides an estimate that indicates how useful each parameter was when building extended decision trees in the model. The more an attribute is used to make key decisions with decision trees, the higher its relative importance.

This importance is explicitly computed for each attribute in the dataset, which makes it possible to rank the attributes and compare them with each other.

For a single decision tree  $T$ , Breiman et al. (1984) proposed as a measure of relevance for each predictor variable  $X_\ell$ :

$$L_\ell^2(T) = \sum_{t=1}^{J-1} \hat{i}_t^2 I(v(t) = \ell) \quad (1)$$

The amount is calculated by  $J - 1$  internal nodes of the tree. In each such node  $t$ , one of the input variables  $X_{v(t)}$  is used to divide the area associated with this node into two subregions; within each individual area correspond to the response values. The selected variable is the one that gives the maximum estimate of an improvement in  $\hat{i}_t^2$  squared risk of error compared to this for a constant fit over the entire region. The square of the relative importance of the variable  $X_\ell$  is the sum of such improvements squared over all internal nodes for which it was chosen as the separation variable. This measure of importance is easily generalized to additive decompositions of trees:

$$L_\ell^2 = \frac{1}{M} \sum_{m=1}^M L_\ell^2(T_m) \quad (2)$$

### 1.3 Principal Component Analysis

Multivariate analysis often works with data that includes a significant number of correlated variables. Principal Component Analysis (PCA) is a size reduction tool that can be used to reduce a large set of variables into a small volume of parameters, which still contains most of the information.

Principal component analysis is a mathematical procedure that converts a series of (possibly) correlated variables into a (smaller) number of uncorrelated variables, called principal components.

The main steps of the algorithm are as follows:

1. Calculate the covariance matrix of X data points.
2. Calculate the eigenvectors and the corresponding eigenvalues.
3. Sort eigenvectors by their eigenvalues in descending order.
4. Select the first k eigenvectors; these will be the new k dimensions.
5. Conversion of initial points of n measurements into k measurements.

The transformation operator is a matrix of eigenvectors sorted by eigenvalues.

#### **1.4 Neural networks**

Artificial neural networks are one of the main tools used in machine learning. In 1943, McCulloch and Pitts proposed a mathematical model of a neuron. This is an algorithm that accepts an n-dimensional vector of an attribute description  $x = (x_1, \dots, X_n)$  as input. The impulses entering the neuron are added to the weights  $w_1, \dots, w_n$ . If the weight is positive, then the corresponding synapse is exciting, if negative, then inhibiting. If the total pulse exceeds the specified activation threshold  $w_0$ , then the neuron is excited and outputs 1, otherwise it is 0. So the neuron calculates the n-ary boolean function:

$$a(x) = \varphi\left(\sum_{j=1}^n w_j x^j - w_0\right) \quad (3)$$

Based on the principle of minimizing empirical risk, the task of tuning synaptic weights can be reduced to finding the vector  $w$ , which delivers a minimum of quality functionality:

$$Q(w) = \sum_{i=1}^l L(a(x_i), y_i) \rightarrow \min_w \quad (4)$$

Using the gradient descent approach and the quality functional, we can write the equation of the iterative update of the neural network weights:

$$w = w - \eta \frac{dQ}{dw} \quad (5)$$

An artificial neural network contains many configuration parameters, including the  $M$  number of neurons and hidden layers of  $H$ . To cover a larger number of values, an additional neuron, called a displacement neuron, is used.

Such networks are taught using an algorithm based on gradient descent called reverse propagation: thanks to the analytical differentiation of superposition while preserving the necessary intermediate values, a large number of operations are required to calculate the gradient, increasing in proportion to the number of weights. It has several advantages:

- High enough efficiency.
- Only information about neurons connected with it passes through each neuron.
- High degree of generality.

Thus, a general view of the ANN is presented in Figure 1.

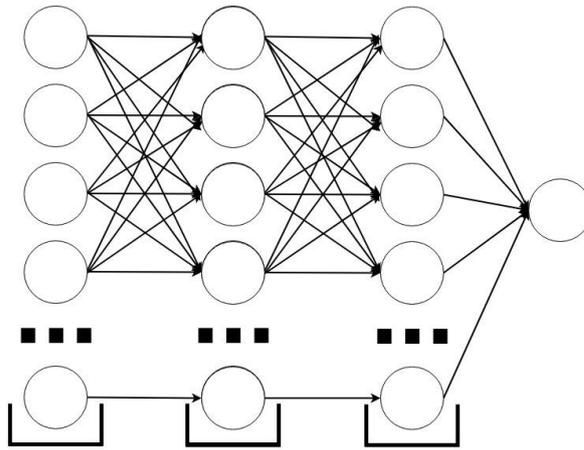


Figure 1- Artificial Neural Network Structure

### 1.4.1 Activation Functions

In addition, each layer of the neural network may contain different activation functions. Most commonly used are:

The sigmoidal activation function is a function of type:

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (6)$$

The range of values is from 0 to 1. This is an S-shaped curve. It is easy to analyze and apply, but it has a number of drawbacks, such as damping of the gradient and low speed of convergence.

Hyperbolch tangent is a function of the form activation:

$$f(x) = \frac{2}{(1 + e^{-2x})} - 1 \quad (7)$$

It is expressed in terms of a sigmoid function, therefore it has the same drawbacks, except for the fact that it is limited to the range (-1, 1).

ReLU - the function gives the output x, if x is positive, and 0 otherwise.

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Despite the fact that with such activation there is a chance of a gradient explosion, it allows you to avoid the attenuation of the gradient. Because of the horizontal line in ReLu (for negative X), the gradient can go to 0. For activations in

this area, the ReLu gradient will be 0, due to which the weights will not be adjusted during the gradient descent. This means that those neurons that enter this state will stop responding to changes in the error / input.

The exponential linear unit — ELU — is a function that, as a rule, reduces losses to zero more quickly and gives more accurate results. It has an output  $x$  if  $x$  is positive and  $\alpha (\exp (x) -1)$  otherwise.

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(e^x - 1) & \text{otherwise} \end{cases} \quad (9)$$

Unlike other activation functions, ELU has an additional alpha constant, which should be a positive number. ELU is very similar to ReLU, except for negative inputs. They are both in the form of an identical function for non-negative input. On the other hand, the ELU is slowly smoothed until its output becomes equal to  $\alpha$ , while the RELU is dramatically smoothed.