

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

Школа инженерного предпринимательства
Направление подготовки 27.03.05 Инноватика

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Применение big–data анализа комплаентности блогеров в digital– маркетинге.

УДК 339.138:005.336.6:004.65.047.44

Студент

Группа	ФИО	Подпись	Дата
ЗН71	Туяков Тамирлан		14.06.21

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кривовяз Н.В.	к.э.н		14.06.21

КОНСУЛЬТАНТЫ:

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Мезенцева И.Л.	–		

Нормоконтроль

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Программист	Долматова А.В.	–		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Корниенко А.А.	к.т.н.		

Томск – 2021

Планируемые результаты освоения ООП

27.03.05 Инноватика

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач
УК(У)-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений
УК(У)-3	Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде
УК(У)-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном и иностранном (-ых) языке
УК(У)-5	Способен воспринимать межкультурное разнообразие общества в социально-историческом, этическом и философском контекстах
УК(У)-6	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни
УК(У)-7	Способен поддерживать должный уровень физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности
УК(У)-8	Способен создавать и поддерживать безопасные условия жизнедеятельности, в том числе при возникновении чрезвычайных ситуаций
УК(У)-9	Способен проявлять предприимчивость в профессиональной деятельности, в т.ч. в рамках разработки коммерчески перспективного продукта на основе научно-технической идеи
Общепрофессиональные компетенции	
ОПК(У)-1	Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности
ОПК(У)-2	Способность использовать инструментальные средства (пакеты прикладных программ) для решения прикладных инженерно-технических и технико-экономических задач, планирования и проведения работ по проекту
ОПК(У)-3	Способность использовать информационно-коммуникационные технологии, управлять информацией с использованием прикладных программ деловой сферы деятельности, использовать компьютерные

	технологии и базы данных, пакеты прикладных программ управления проектами
ОПК(У)-4	Способность обосновывать принятие технического решения при разработке проекта, выбирать технические средства и технологии, в том числе с учетом экологических последствий их применения
ОПК(У)-5	Способность использовать правила техники безопасности, производственной санитарии, пожарной безопасности и нормы охраны труда
ОПК(У)-6	Способность к работе в коллективе, организации работы малых коллективов (команды) исполнителей
ОПК(У)-7	Способность применять знания математики, физики и естествознания, химии и материаловедения, теории управления и информационные технологии в инновационной деятельности
ОПК(У)-8	Способность применять знания истории, философии, иностранного языка, экономической теории, русского языка делового общения для организации инновационных процессов
Профессиональные компетенции	
ПК(У)-1	Способность использовать нормативные документы по качеству, стандартизации в практической деятельности
ПК(У)-2	Способность использовать инструментальные средства (пакеты прикладных программ) для решения прикладных инженерно-технических и технико-экономических задач, планирования и проведения работ по проекту
ПК(У)-3	Способность использовать информационно-коммуникационные технологии, управлять информацией с использованием прикладных программ деловой сферы деятельности; использовать сетевые компьютерные технологии и базы данных в своей предметной области, пакеты прикладных программ для анализа, разработки и управления проектом
ПК(У)-4	Способность анализировать проект (инновацию) как объект управления
ПК(У)-5	Способность определять стоимостную оценку основных ресурсов и затрат по реализации проекта
ПК(У)-6	Способность организовать работу исполнителей, находить и принимать управленческие решения в области организации работ по проекту и нормированию труда
ПК(У)-7	Способность систематизировать и обобщать информацию по использованию и формированию ресурсов
ПК(У)-8	Способность применять конвергентные и мультидисциплинарные знания, современные методы исследования и моделирования проекта с использованием вычислительной техники и соответствующих программных комплексов

ПК(У)-9	Способность использовать когнитивный подход и воспринимать (обобщать) научно-техническую информацию, отечественный и зарубежный опыт по тематике исследования
ПК(У)-10	Способность спланировать необходимый эксперимент, получить адекватную модель и исследовать ее
ПК(У)-11	Способность готовить презентации, научно-технические отчеты по результатам выполненной работы, оформлять результаты исследований в виде статей и докладов
ПК(У)-12	Способность разрабатывать проекты реализации инноваций с использованием теории решения инженерных задач и других теорий поиска нестандартных, креативных решений, формулировать техническое задание, использовать средства автоматизации при проектировании и подготовке производства, составлять комплект документов по проекту
ПК(У)-13	Способность использовать информационные технологии и инструментальные средства при разработке проектов
ПК(У)-14	Способность разрабатывать компьютерные модели исследуемых процессов и систем
ПК(У)-15	Способность конструктивного мышления, применять методы анализа вариантов проектных, конструкторских и технологических решений для выбора оптимального
ПК(У)-16	Способность выполнения работ по сопровождению информационного обеспечения и систем управления проектами
ПК(У)-17	Способность ведения баз данных и документации по проекту
Профессиональные компетенции университета	
ДПК(У)-1	Способность к экономическому планированию деятельности структурного подразделения промышленной организации, которое направлено на организацию рациональных бизнес-процессов в соответствии с потребностями рынка, обеспечение участия работников структурного подразделения промышленной организации в проведении маркетинговых исследований

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

Школа инженерного предпринимательства
 Направление подготовки /27.03.05 «Инноватика»

УТВЕРЖДАЮ:
 Руководитель ООП
 _____ А.А. Корниенко
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы	
Студенту:	
Группа	ФИО
3Н71	Туякову Тамирлану
Тема работы:	
Применение big-data анализа комплаентности блогеров в digital- маркетинге.	
Утверждена приказом директора (дата, номер)	№112-10/с от 22.04.2021

Срок сдачи студентом выполненной работы:	14.06.21
--	----------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

<p>Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<ol style="list-style-type: none"> 1. Электронные ресурсы 2. Статьи в журналах 3. Материалы отчета по УИРС студента
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> 1. Аналитический обзор реализации передачи данных 2. Объект и методы исследования 3. Реализация серверной и клиентской сторон

<p>Перечень графического материала (с точным указанием обязательных чертежей)</p>	<p>Рисунок 1 – Общий вид нейронной сети Рисунок 2 – динамика показателей аккаунта блогера за год Рисунок 3 – Общий вид данных Рисунок 4 – Распределение данных Рисунок 5 – Общий вид нейронной сети Рисунок 6 – График процесса обучения Рисунок 7 – Процесса обучения Рисунок 8 – ReLU_2 слоя_48 нейронов Рисунок 9 – ELU_2 слоя_54 нейронов Рисунок 10 – Tanh_3 слоя_30 нейронов Рисунок 11 – Функция принадлежности переменной K Рисунок 12 – Функция принадлежности переменной Sh Таблица 1 – Популярные библиотеки машинного обучения Таблица 2 – Основные параметры Нейронной сети Таблица 3 – Оптимальные параметры Нейронной сети Таблица 4 – Работа нечёткой логики Таблица 5 – Сопоставление K и доход от продажи курсов по доп. Таблица 6 – Карта сегментирования рынка прогнозирующих систем Таблица 7 – Оценочная карта для сравнения конкурентных программных решений Таблица 8 – Морфологическая матрица проекта Таблица 9 – Сводная информация о заинтересованных сторонах Таблица 10 – Сводная информация о целях проекта Таблица 11 – Организационная структура Таблица 12 – Сводная информация о факторах проекта Таблица 13 – Структура работы проекта Таблица 14 – Группировка затрат по статьям Таблица 15 – Расчет материальных трат Таблица 16 – Расчет технических трат Таблица 17 – Баланс рабочего времени Таблица 18 – Расчёт основной заработной платы Таблица 19 – Отчисления на социальные нужды</p>
--	---

	<p>Таблица 20 – Оценка рисков проекта</p> <p>Таблица 21 – Разница затрат по каждому исполнению НТИ</p> <p>Таблица 22 – Расчет интегральных показателей ресурсоэффективности</p> <p>Таблица 23 – Сравнительная эффективность разработок</p> <p>Таблица 24 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений</p> <p>Приложение А (обязательное) – Матрица SWOT –анализа</p> <p>Приложение Б (обязательное) – Степень компетенций в реализации проекта</p> <p>Приложение В (обязательное) – вредные и опасные факторы, возникающие при работе с ПЭВМ</p>
--	--

Консультанты по разделам выпускной квалификационной работы
(с указанием разделов)

Раздел	Консультант
Нормоконтроль	Долматова А.В.
Социальная ответственность	Мезенцева И.Л.
Названия разделов, которые должны быть написаны на русском и иностранном языках:	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику

--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кривовяз Н.В.	к.э.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
5Н71	Туяков Тамирлан		

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

Школа инженерного предпринимательства
Направление подготовки 27.03.05 Инноватика
Уровень образования бакалавриат
Период выполнения (осенний / весенний семестр 2020/2021 учебного года)

Форма представления работы:

Бакалаврская работа

Применение big–data анализа комплаентности блогеров в digital– маркетинге
--

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	14.06.2021
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
01.03.21	Определение темы ВКР и получение задания.	
03.03.21	Согласование плана ВКР с руководителем. Составление окончательного плана работы.	
19.03.21	Сбор и проработка материала для написания первой главы. Отправка главы руководителю на проверку.	
09.04.21	Сбор и проработка материала для написания второй главы. Отправка главы руководителю на проверку.	
28.04.21	Сбор и проработка материала для написания третьей главы. Отправка главы руководителю на проверку.	
19.05.21	Сбор и проработка фактического материала для написания главы «Социальная ответственность». Отправка главы на проверку консультанту по разделу.	
28.05.21	Исправление замечаний консультанта по разделу «Социальная ответственность»	
12.06.21	Исправление замечаний и доработка ВКР, оформление согласно стандартам, отправка готовой работы руководителю на проверку.	

13.06.21	Подготовка презентации для защиты ВКР и доклада.	
----------	--	--

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кривовяз Н. В	к.э.н.		

Принял студент:

ФИО	Подпись	Дата
Туяков Тамирлан		

СОГЛАСОВАНО:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Корниенко А.А	к.т.н.		

Реферат

Выпускная квалификационная работа содержит 84 страницы, 12 рисунков, 24 таблицы, 35 литературных источников, 3 приложения.

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 и создание проекта прогноза комплаентности блогера.

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.....	10
Введение	12
1 Теоретические основы создания data-mining алгоритмов.....	14
1.1 Выбор программных компонентов	14
1.2 Разведочный анализ данных	16
1.3 Разложение на главные компоненты	19
1.4 Подготовка данных.....	20
1.5 Нейронные сети.....	22
2 Социальные сети как поле влияния.....	26
2.1 Аналитический обзор социальных сетей.....	26
2.2 Аналитический обзор инструментов сегментации аудитории.....	30
2.3 Оценка влияния блогера на аудиторию. Процедура оценки комплаентности.....	32
3 Формирование моделей прогнозирования комплаентности аудитории.....	38
3.1 Нейронная сеть.....	38
3.2 Проектирование нечёткой логики	42
4 Оценка потенциала коммерциализации алгоритма скоринга блогеров	45
5 Социальная ответственность	63
Заключение	73
Список использованных источников	75
Приложение А Матрица SWOT – анализа	80
Приложение Б Степень компетенций в реализации проекта	82
Приложение В Вредные и опасные факторы, возникающие при работе с ПЭВМ.....	84

Введение

На сегодняшний день ведение блога — популярный и удобный способ для публикации новостей, распространения мнений и формирования виртуальных сообществ.

Степенной закон всех блогов заключается в том, что чрезвычайно влиятельных блогов очень мало. Такие блоги могут формировать общественное мнение и даже с подвигать аудиторию совершать поступки разного рода. 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. Прежде чем люди покупают или принимают решения, они слушают чужой опыт, мнения и предположения.

Маркетологи разных бизнесов ищут именно влиятельных блогеров для рекламных размещений в их публикациях и контенте. Однако, спрогнозировать возможный результат от размещения очень трудно. Это причина по которой в агентства digital-маркетинга в последние 4 года активно поступают запросы на PR коллаборации с блогерами и ORM макркетинг. И даже рекламные агентства допускают ошибки и коллаборации проходят неуспешно.

Объектом исследования в выпускной квалификационной работе выступает сервис анализа просчета уровня комплаентности аудитории блогера для digital-маркетологов. Предметами исследования выступают данные из открытых аналитических источников о вовлеченности и базовой аналитике контента блогера и взаимодействиях аудитории с этим контентом.

Цель работы заключается в доказательстве технической реализации сервиса по расчету комплаентности блогера на основе big-data анализа вовлеченности его аудитории.

Для достижения поставленной цели необходимо выполнить ряд задач:

- Провести анализ теоретического материала создания data-mining алгоритмов;
- Разработать математическую модель подсчета комплаентности блогера;
- Провести расчет потенциала коммерциализации алгоритма скоринга блогеров.

В ходе работы было проведено исследование и анализ программных компонентнов, на основе которых можно рассчитать, согласно математической модели проекта, комплаентность конкретного блогера.

Область применения: результаты проведенных исследований могут быть использованы для вывода и реализации нового продукта на рынок онлайн сервисов для интернет маркетологов.

Результат ВКР имеет практическую значимость для маркетологов ООО «ТД СОЛО ИК». Практическая значимость заключается в возможности использования прогнозирования результатов размещения рекламной публикации у блогера.

1 Теоретические основы создания data-mining алгоритмов

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/versions/r0.12/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/versions/r0.12/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.

Отсутствие значений параметра

Пропущенные значения отражают беспорядок в данных реального мира. Причины 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.

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.

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 .

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_ℓ следующая формула:

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

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_ℓ 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:

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

1.3 Разложение на главные компоненты

Многофакторный анализ часто работает с данными, включающими значительное количество коррелированных переменных. Анализ главных компонент (PCA) 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

Для использования data mining алгоритмов предварительно нужно подготовить набор анализируемых данных. Такие данные должны быть в достаточном объеме, чтобы репрезентативно определить закономерность в наборе данных. Также данные должны быть приемлемо объемны, так как обработка большого объема данных будет занимать большое количество времени.

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.

Целевая функция определяется в зависимости от цели анализа. Выбор нужной функции имеет ключевое значение для успешного 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.

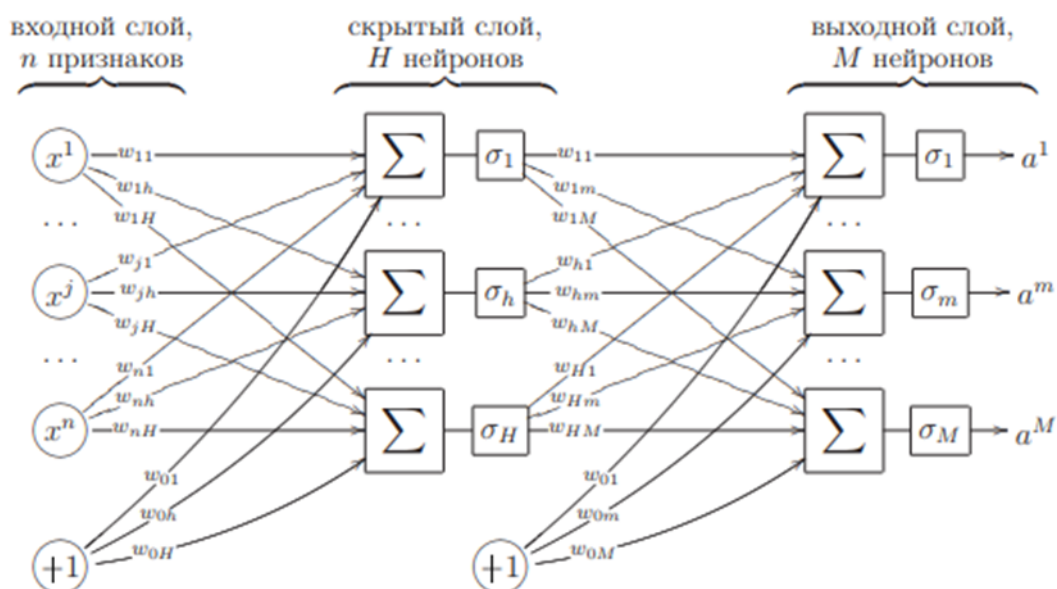


Рисунок 1 – Общий вид нейронной сети

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 ограничен 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].

2 Социальные сети как поле влияния

2.1 Аналитический обзор социальных сетей

«Социальная сеть» — ресурс, на котором реализована возможность передачи сообщений между пользователями. Также, помимо обмена сообщениями подобные ресурсы дают возможность публиковать контент, смотреть и реагировать на контент других пользователей социальной сети.

Ресурсы такого типа пользуются большой популярностью в интернете. Особенно если учитывать сторону бизнес-проектов. Среди пользователей таких ресурсов находится много целевой аудитории обычных бизнесов. Поэтому, последние используют разные маркетинговые стратегии продвижения компании в социальных медиа.

Социальные сети для общения (Relationship networks):

Самый популярный тип социальных медиа — это ресурсы для обмена сообщениями. Самая популярная сеть в Российской Федерации — ВКонтакте, как раз яркий представитель социального медиа такого типа.

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

Подвид социальных медиа, в котором упор стоит на потреблении фото- или видеоматериала.

Социальные сети для авторских записей (Social publishing platforms)

Социальная сеть или сеть микроблогов — Twitter, как яркий пример ресурса авторских записей.

Основные причины популярности социальных медиа

1. Облегченные коммуникативные способности. Простота в обращении с ресурсами в моменте передачи сообщений. Возможность взаимодействия с медийными личностями.

2. Повышение чувства собственной важности — измеримое в количество реакций под публикациями. Пользователи таких ресурсов радуются и разочаровываются от количества реакций под их публикациями.

3. Бизнес задачи — реализация своих продуктов, реклама и выстраивание доверия к бренду с помощью маркетинговых инструментов социальных медиа — вошло в норму. Отзывы пользователей, взаимодействия и коллаборации с блогерами вызывают доверие к бренду и рост продаж бизнеса.

4. Скорость работы и доступность. На данный момент интернет доступен практически в любом населенном месте планеты. В связи с этим облегчился доступ в социальные медиа.

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

Аудитории социальных сетей

Тенденции и перспективы

Рецензия We Are Social и Hootsuite наглядно показывает большой рост юзеров социальных медиа. Каждый второй человек в мире зарегистрирован в одном из представителей социальных медиа. Более 1 миллиона регистраций проходят ежедневно. При этом основной рост заключен в пользователях мобильных девайсов — практически 80% пользователей используют мобильный девайс при входе в социальную сеть.

Также есть тренд на падение уровня доверия пользователей к медиа и к публикациям внутри социальных медиа. Этот тренд возник на основе перманентного роста количества рекламных публикаций. На данный момент, почти каждый бизнес имеет представительство в социальных сетях.

Начинающие IT предприниматели, создающие новые конкурентные решения на рынке социальных медиа, привлекают и обращают на себя внимание аудитории уже устоявшихся социальных сетей. В связи с этим,

каждая социальная сеть разрабатывает маркетинговую кампанию с глубоким анализом всех конкурентов на рынке и с расчетом тенденций рынка.

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

Главная социальная сеть в мире. Главная площадка по закупке размещения рекламы. Сервис общения, который стал эталонным для многих последующих социальных сетей. Сервис дает возможность:

1. Создать страницу; выкладывать посты с фото-видео содержанием.
2. Совершать поиск пользователей внутри сети. Общаться и обмениваться информацией с другими пользователями.
3. Делиться материалами и выкладывать их общедоступно.
4. Создавать группы и объединять в них пользователей по интересам.
5. Читать и просматривать новостную ленту.
6. Комментировать посты и ставить реакции.

Социальная сеть перманентно вводит новшества в продукт. Компания является с самыми передовыми возможностями таргетирования рекламы. Это позволяет удерживать рекламодателей. При этом Facebook не забывает про обычных пользователей и вводит новшества и для них.

Twitter

Сеть микроблогов. Социальная сеть взяла на себя роль трендсеттера и главной социальной сети по скорости публикации новостей и создания инфоповодов.

Ключевая функция: ведение аккаунта малыми постами (твитами) с емкостью в 280 знаков. Есть расширенный алгоритм подбора новостей,

который показывает пользователю интересную новость вне зависимости от подписки на публикации автора.

LinkedIn

Чаще всего эта социальная сеть используется для рекрутинга и нетворкинга в профессиональной среде. Ставка сделана на профессиональном контенте, интересным определенному кругу пользователей.

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

Социальная сеть, заточенная под мобильные устройства и контент с сильным визуальным составляющим. На данный момент у этого ресурса более 1 миллиарда пользователей по всему миру.

Функции:

1. Ведение фотоблогов.
2. Обработка фотографий – эффекты, рамки, специальные режимы съемки, коллажи.
3. Создание своего профиля и ленты публикаций.
4. Скрининг аккаунтов других пользователей, оценка и комментирование их публикаций.
5. Shopping-tags — создание каталога продуктов бизнеса.
6. Поиск нужной информации по хештегам (специальным отметкам).

Instagram – одна из самых популярных социальных сетей в мире. Сильная визуальная составляющая контента дало Instagram выгодное конкурентное преимущество.

ВКонтакте

Социальная сеть вдохновленная успехом Facebook. Самая популярная социальная сеть на территории Российской Федерации.

Возможности:

1. Создание аккаунта.

2. Передача сообщений между пользователями.

3. Функция файлообменника — возможность пересылать другим пользователям фото- и видеоматериалы. А также другие файлы.

4. Онлайн-гейминг на базе Mail Ru Group.

YouTube

Видеохостинг YouTube — главная платформа по публикации видеоматериала. Является социальной сетью, так как есть возможность передачи сообщений между пользователями внутри хостинга.

Возможности платформы:

1. Выкладывать видеоматериал в открытый либо ограниченный доступ на платформе.

2. Редактура видеоматериала при загрузке. Выбор заставки.

3. Комментировать, оценивать чужие видео или комментарии.

4. Ведение лайв эфиров.

YouTube активно борется за цензуру внутри своего продукта. В рамках этой борьбы YouTube отключает монетизацию видео либо вовсе удаляет видео с платформы. Примечательно, что на YouTube большое давление со стороны властей разных стран, которые требуют удалить либо дать персональную информацию об авторе определенного ролика.

2.2 Аналитический обзор инструментов сегментации аудитории

Сегментация — это выделение подаудитории внутри целевой аудитории бизнеса. Каждая подаудитория объединена по признаку схожести. Например, по полу. Сегментация необходима в digital-маркетинге, так как внутренних инструментов для продвижения бренда на ресурсах социальных медиа слишком много. Соответственно аудитория может быть с разными предпосылками к покупке продукта. В таком случае сегментация поможет увеличить рентабельность инвестиций в маркетинг.

Инструменты сегментирования аудитории дробят аудиторию на группы пользователей, а внутренние инструменты рекламных площадок помогают показать рекламу каждой группе с релевантным маркетинговым посылом к покупке.

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 сетях
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.

2.3 Оценка влияния блогера на аудиторию. Процедура оценки комплаентности

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 пользователи.

Показатель «вовлеченность» как раз и измеряет количество действий в зависимости от количества аудитории аккаунта. Число подписчиков уже давно не является основополагающим фактором при взаимодействии с блогерами.

Сервисы по аналитике профиля в Инстаграм, такие как EpicDetect или LiveDune, используют коэффициент вовлеченности ER. Рассчитывают его по формуле:

$$ER = (\text{лайки} + \text{комментарии}) / \text{подписчики} * 100\% \quad (10)$$

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.

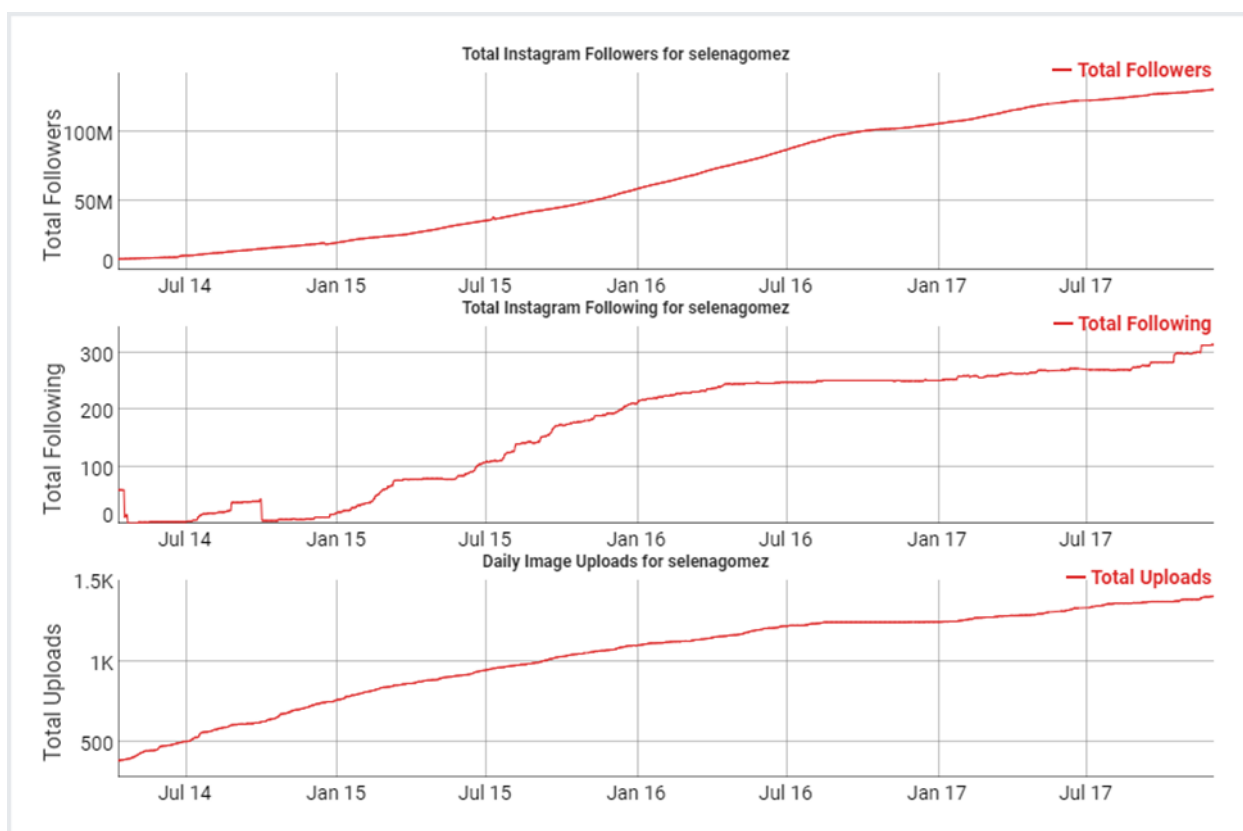


Рисунок 2 – динамика показателей аккаунта блогера за год

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.

Предобработка данных

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 совокупности: 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.

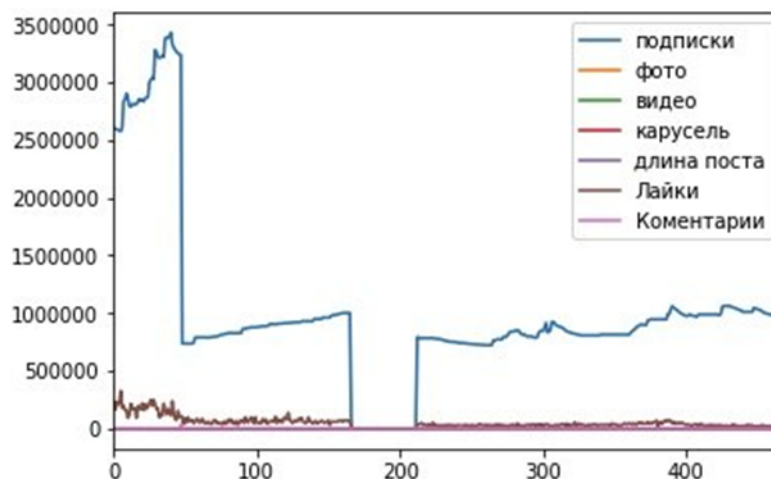


Рисунок 4 – Распределение данных

На рисунке 4 видны все операции по проверке данных.

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 блоггер.

При этом итоговая комплаентность (K):

$$K(Bloger_i) = \frac{f(N_i) * ER_i * \sum_{j \in Z_i} I_{\{T_j\}} * a_i}{\sum_{j \in \{1,2,\dots,M\}} (\sum_{k \in Z_j} I_{\{T_k\}} * a_k) * ER_j * f(N_j)}, \quad i=1,2,3,\dots,M \quad (11)$$

Формула 11. – комплаентность аудитории блогера

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

Таким образом Формула 11 отображает зависимость комплаентности аудитории от заданных параметров.

3 Формирование моделей прогнозирования комплаентности аудитории

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.

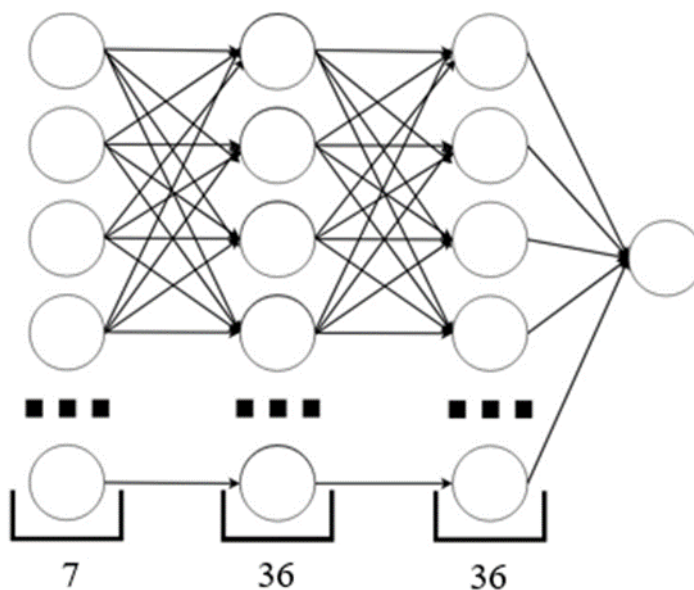


Рисунок 5 – Общий вид нейронной сети

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

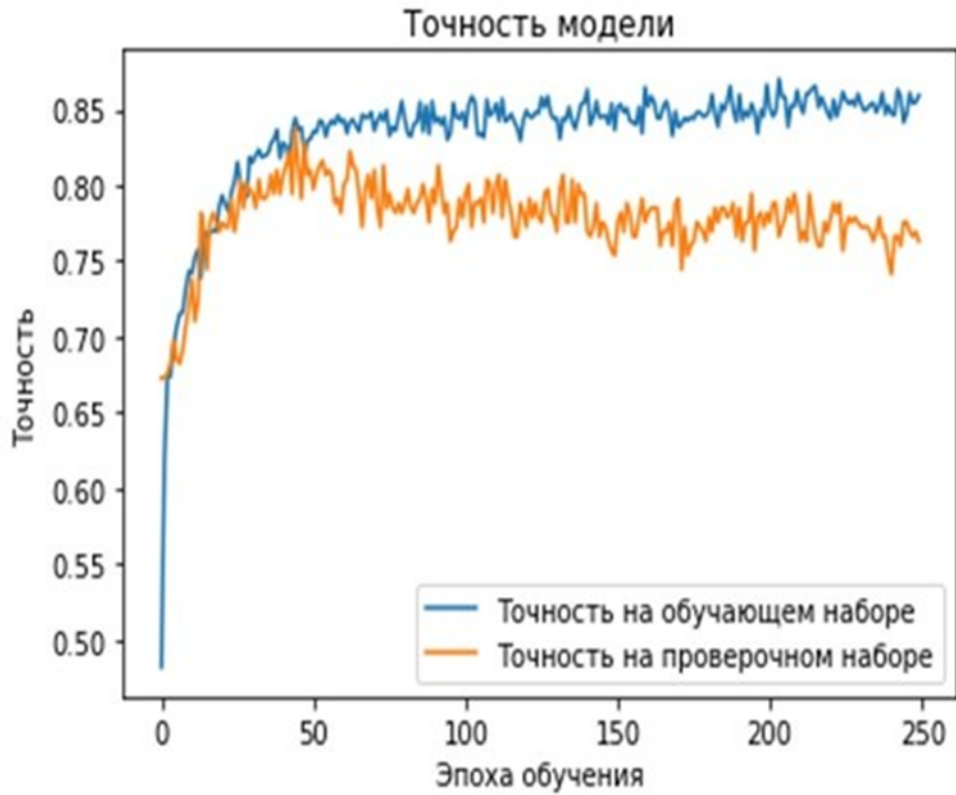


Рисунок 6 – График процесса обучения

```

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: 0.9104
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: 0.9104
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: 0.9104
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: 0.9104
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: 0.9104
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: 0.9104
Epoch 1000/1000
469/469 [=====] - 0s 51us/sample - loss: 35.6800 - acc: 0.9104

```

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

```
Epoch 995/1000
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 Проектирование нечёткой логики

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

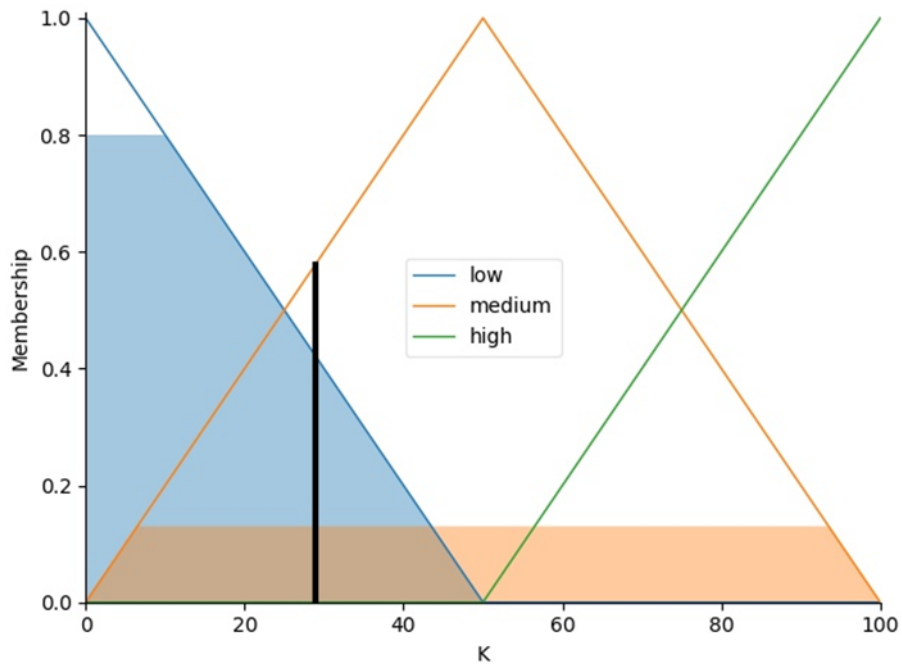


Рисунок 11 – Функция принадлежности переменной К

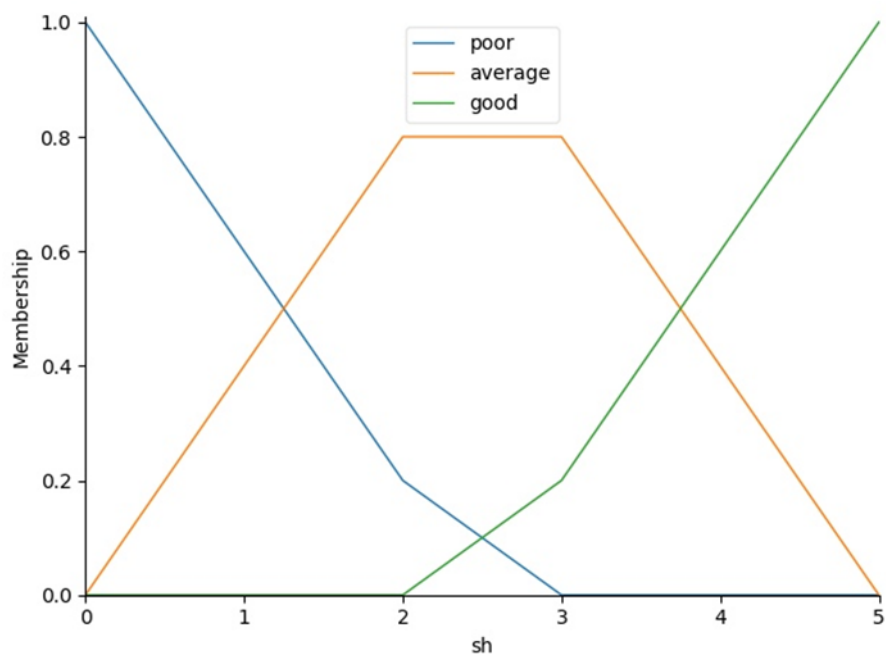


Рисунок 12 – Функция принадлежности переменной Sh

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 Оценка потенциала коммерциализации алгоритма скоринга блогеров

Конечным продуктом 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.

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. Вторым критерий – размеры компании–заказчика.

В таблице 6 представлена карта сегментирования рынка на основе наиболее значимых критериев.

Таблица 6 – Карта сегментирования рынка прогнозирующих систем:

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

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.

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 7.

Таблица 7 – Оценочная карта для сравнения конкурентных программных решений

Критерии оценки	Вес критерия	Баллы				Конкурентноспособность			
		Бф	Бк 1	Бк 2	Бк 3	Кф	Кк 1	Кк 2	Кк3
Технические критерии оценки ресурсоэффективности									
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

Продолжение таблицы 7

Уровень защиты доступа	0,1	5	5	5	5	0,5	0,5	0,5	0,5
Простота интерфейса	0,1 5	3	3	2	3	0,4 5	0,4 5	0,3	0,45
Экономические критерии оценки ресурсоэффективности									
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	0,4 5	0,6	0,45
Уровень применения на рынок	0,0 5	0	3	3	4	0	0,1 5	0,1 5	0,2
Поддержка продукта	0,05	1	3	3	5	0,0 5	0,1 5	0,1 5	0,25
Итого	1					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. Специализированный функционал подразумевает сфокусированность на прогнозе комплаентности и машинном обучении модели.

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, которые указаны в Приложении А.

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, которая указана в Приложении Б.

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.

Коммерциализация результатов научно–технического исследования

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.

Инициация проекта

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

Определение возможных альтернатив исследований

Для определения возможных альтернативных путей проведения 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 слоев	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.

Задачи 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.

Таблица 9 – Сводная информация о заинтересованных сторонах

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

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

Таблица 10 – Сводная информация о целях проекта

Цели проекта:	Разработка алгоритмов анализа популярности блогера у его целевой аудитории в социальных сетях
---------------	---

Продолжение таблицы 10

Ожидаемые результаты проекта	Программа, способная количественно оценить популярность блогера у его аудитории.
Критерии приемки результата проекта:	Программа удовлетворяет критериям точности, надежности, быстродействия.
Требования к результату проекта:	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.)

Организационная структура проекта

Таблица 11 – Организационная структура

№ п/п	ФИО, основное место работы, должность	Роль в проекте	Функции	Трудовые затраты, час.
1	Туяков Т, студент	Руководитель	Координация действий и решений, Проверка реализации	300
2	Погожев А.О., технический директор	Разработчик	Разработка программной системы	381
ИТОГО:				681

Ключевой вопрос заключается в составе рабочей группы проекта. Также необходимо определить роли участников и прописать выполняемые функции с трудозатратами в проекте.

Ограничения и допущения проекта

Ограничения проекта – это все факторы, которые могут послужить ограничением степени свободы участников команды проекта, а также «границы проекта» – параметры проекта или его продукта, которые не будут реализованных в рамках данного проекта

Таблица 12 – Сводная информация о факторах проекта

Фактор	Ограничения/ допущения
Бюджет проекта	200 000 руб.
Источник финансирования	ООО «ТД СОЛО ИК»
Сроки проекта:	4 месяца
Дата утверждения плана управления проектом	02.02.2021
Дата завершения проекта	02.06.2021
Прочие ограничения и допущения*	Отсутствуют

Структура работ в рамках 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.

Таблица 13 – Структура работы проекта

Код работы	Название	Д н и	Дата начала работ	Дата окончания работ	Состав участников
1	Составление ТЗ	2	02.02.2021	04.02.2021	Бакалавр, Руководитель
2	Поиск и исследование литературы	10	05.02.2021	15.02.2021	Бакалавр
3	Выбор направления исследования	10	16.02.2021	27.02.2021	Бакалавр, Руководитель
4	Теоретические исследования и расчёты	15	28.02.2021	18.03.2021	Бакалавр, Руководитель

Продолжение таблицы 13

5	Разработка алгоритмов предсказания данных	22	19.03.2021	12.04.2021	Бакалавр
6	Тестирование разработанных алгоритмов	14	13.04.2021	29.04.2021	Бакалавр
7	Оформление отчета НИР	27	30.04.2021	01.06.2021	Бакалавр, Руководитель
Итого:		100	02.02.2021	01.06.2021	

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

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 14 – Группировка затрат по статьям

Вид работ	Статьи									Итого плановая себестоимость
	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 предприятиями	Прочие прямые расходы	Накладные расходы	
1.	566,5	29900	92693,55	9 269, 36	25119,95	0	3200	174	30588,87	191512,23

Расчет материальных затрат

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).

Таблица 15 – Расчет материальных трат

Наименования	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

Расчет технических затрат: специальное оборудование

Определение стоимости спецоборудования производится по действующим прейскурантам, а в ряде случаев по договорной цене.

Таблица 16 – Расчет технических трат

№ 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

Расчет заработной платы

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

$$C_{зп} = Z_{осн} + Z_{доп} \quad (12)$$

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

$Z_{доп}$ – дополнительная заработная плата.

Основная заработная плата ($Z_{осн}$)

$$Z_{осн} = Z_{дн} \cdot T_{раб} \quad (13)$$

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

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

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{дн} = \frac{Z_m \cdot M}{F_d} \quad (14)$$

Где Z_m – 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_d – действительный годовой фонд рабочего времени научно-технического персонала, раб.дн.

Месячный должностной оклад работника:

$$Z_m = Z_b \cdot (k_{пр} + k_d) \cdot k_r \quad (15)$$

где Z_b – базовый оклад, руб.;

$k_{пр}$ – премиальный коэффициент;

k_d – коэффициент доплат и надбавок;

k_r – 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 17 – Баланс рабочего времени

Показатели рабочего времени	Руководитель	Разработчик
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 предприятия.

Таблица 18 – Расчёт основной заработной платы

Исполнители	Зб, руб.	кр	Зм, руб.	Здн, руб.	Тр, раб.дн.	Зосн, руб.
Руководитель	33 664	1,3	43 763,2	1 813,3	13,5	24 479,55
Разработчик	12 664	1,3	16 463,2	682,14	100	68 214
Итого						92 693,55

$$Z_{осн} = 92\,693,55 \text{ руб.}$$

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

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

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 19 перечислены отчисления на каждого из работников.

Таблица 19 – Отчисления на социальные нужды

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

$$C_{внеб} = 25\,119,95 \text{ руб}$$

Накладные расходы

$$C_{накл} = (Z_{доп} + Z_{осн}) * 30\% \quad (16)$$

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

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:

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

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:

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

Итоговые затраты

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

$$C_{\text{проекта}} = C_{\text{м}} + Z_{\text{осн}} + Z_{\text{доп}} + C_{\text{внеб}} + C_{\text{о}} + C_{\text{контр}} + C_{\text{электр}} + C_{\text{накл}} \quad (17)$$

$$C_{\text{проекта}} = 566,5 + 92\,693,55 + 9\,269,36 + 25\,119,95 + 29\,900 + 3200 + 174 + 30\,588,87$$

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

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

Таблица 20 – Оценка рисков проекта

№	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 в 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 реализации новых подходов	Создание аналогового продукта другими разработчиками

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

Разница затрат по каждому исполнению НТИ представлена в таблице 21.

Таблица 21 – Разница затрат по каждому исполнению НТИ

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

Интегральный показатель ресурсоэффективности можно определить следующим образом:

$$I_{p,i} = \sum a_i b_i \quad (18)$$

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

a_i – весовой коэффициент i -го варианта разработки,

b_i – балльная оценка i -го варианта исполнения разработки,

устанавливаемая экспертным путем по выбранной шкале оценивания,

n – число параметров сравнения.

Расчет интегральных показателей ресурсоэффективности приведен в таблице 22:

Таблица 22 – 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 таблице 23:

Таблица 23 – Сравнительная эффективность разработок

Показатели	Исп1	Исп2	Исп3
Интегральный финансовый показатель разработки $I_{финр}$	0,748	0,988	1

Продолжение таблицы 23

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.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
3Н71	Туякову Тамирлану

Школа	Инженерного предпринимательства	Отделение (НОЦ)	
Уровень образования	Бакалавриат	Направление/специальность	27.03.05 «Инноватика»

Тема ВКР:

Применение big–data анализа комплаентности блогеров в digital– маркетинге.	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования: Сервис анализа комплаентности блогеров для digital–маркетинга Область применения: рынок разработки программного обеспечения для digital–маркетинга
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	<ul style="list-style-type: none"> – ГОСТ 12.2.032–78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования; – ГОСТ Р ИСО 9241–4–2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT). Часть 4. Требования к клавиатуре – ГОСТ Р ИСО 9241–5–2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT). Часть 5. Требования к расположению рабочей станции и осанке оператора.
2. Производственная безопасность: <ul style="list-style-type: none"> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия 	Вредные факторы: <ul style="list-style-type: none"> – Отклонение показателей микроклимата в закрытом помещении; – Отсутствие или недостаток естественного освещения; – Повышенный уровень шума. Опасные факторы: <ul style="list-style-type: none"> – Напряжение в электрической сети, замыкание.

3. Экологическая безопасность:	Атмосфера: утилизация неисправной электроники, утилизация макулатуры
4. Безопасность в чрезвычайных ситуациях:	Возможные ЧС: природные – пожар, аварии на коммунальных системах, обрушение зданий. Наиболее типичная ЧС: пожар по причине работы за ПК.

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

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Мезенцева Ирина Леонидовна	—		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
3Н71	Туяков Тамирлан		

5 Социальная ответственность

Введение

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 невозполнимых ресурсов и защите в чрезвычайных ситуациях.

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.

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.

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.

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.

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.

Производственная безопасность

Норма площади рабочего места с персональным компьютером составляет 4,5 м². В офисе, где выполнялась выпускная квалификационная работа студента, установлено 4 рабочих места с персональными компьютерами и жидкокристаллическими экранами. Соответственно, на одного человека приходится 5 м², что соответствует вышеуказанным требованиям.

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 Приложении В представлены возможные вредные и опасные факторы, возникающие при работе за ПЭВМ.

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.

Недостаточная освещенность рабочей зоны

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

Недостаточная освещенность рабочих мест может являться причиной снижения производительности и качества работы, получения производственных травм. Поэтому качественный свет – залог безопасной работы. Он повышает трудоспособность и снижает риск травматизма на рабочем месте.

Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 – 500 лк (СП 52.13330.2016), Освещенность поверхности экрана не должна быть более 300 лк.

Согласно СП 52.13330.2016 нормы на освещение для оператора поста управления берутся для производственных помещений.

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).

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.

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 24.

Таблица 24 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений

Период года	Категория работы по уровням энергозатрат, Вт	Темп. воздуха, °С	Темп. поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Оптимальные показатели					
Холодный	Iб (140 – 174)	21–23	20–24	60–40	0,1
Теплый	Iб (140 – 174)	22–24	21–25	60–40	0,1
Допустимые показатели					
Холодный	Iб (140 – 174)	19–24	18–25	15–75	0,1–0,2
Теплый	Iб (140 – 174)	20–28	19–29	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.

Повышенный уровень шума на рабочем месте

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

Шум создает значительную нагрузку на нервную систему человека, оказывая на него психологическое воздействие.

Во избежание негативных последствий от производственного шума, его необходимо регулировать в соответствие с нормами, которые указаны в ГОСТ 12.1.003–2014 «Система стандартов безопасности труда. Шум. Общие требования безопасности».

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.

Допустимый уровень звукового давления, согласно ГОСТ 12.1.003–2014, колеблется от 38 дБ до 86 дБ при частоте от 8000 Гц до 31,5 Гц, соответственно.

Для снижения уровня шума, производимого персональными компьютерами, рекомендуется регулярно проводить их техническое обслуживание: чистка от пыли, замена смазывающих веществ; также применяются звукопоглощающие материалы.

Опасные производственные факторы

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

Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека

Электричество, используемое для питания компьютерной техники, может стать источником опасности, а так как сотрудник отдела технического контроля большинство своего рабочего времени проводит за компьютером, он находится в зоне риска.

Поражение электрическим током может произойти при прикосновении к токоведущим частям, находящимся под напряжением, на которых остался заряд или появилось напряжение. Электрический ток оказывает на человека термическое, электролитическое, биологическое и механическое воздействие.

Действие электрического тока на человека приводит к травмам или гибели людей.

Для минимизации опасности воздействия электричества на тело человека необходимо соблюдать правила ГОСТ 12.1.038–82 [10]. Для переменного тока частотой 50 Гц допустимое значение напряжения прикосновения составляет 2 В, а силы тока – 0,3 мА, для тока частотой 400 Гц соответственно – 2 В и 0,4 мА; для постоянного тока – 8 В и 1 мА. Возможные меры защиты: оградительные устройства, устройства автоматического контроля и сигнализации, изолирующие устройства и покрытия, устройства защитного заземления и зануления, устройства автоматического отключения; предохранительные устройства.

Экологическая безопасность

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.

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.

Безопасность в чрезвычайных ситуациях

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».

Выводы по разделу

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 текущих реалиях, практически каждый предприниматель или маркетолог задумывается о рекламном размещении в контенте блогера. Однако практически каждый сталкивается с проблемой выбора конкретного блогера из большого множества. Существующие решения помогают косвенно решить проблему и зачастую вводят неподготовленного пользователя в заблуждение.

На данный момент на рынке digital-маркетинга нет ни одного сервиса способного прогнозировать результат размещения рекламы у блогеров. Рекламные агентства; гос. Корпорации; PR агентства; средние, мелкие и крупные бизнесы закупают рекламные размещения у блогеров и делают коллаборации с ними. Также есть большой спрос на амбассадоров бренда.

В связи с этим у ООО «ТД СОЛО ИК» возникло желание организовать разработку проекта способного автоматически просчитывать комплаентность блогеров. При этом термин «комплаентность» был перенесен из фармацевтического маркетинга.

В результате анализа data-mining алгоритмов и социальных сетей был выявлен оптимальный способ работы с данными через API (Python + TensorFlow) и необходимые параметры на основе которых строится расчет комплаентности аудитории. Основным источником данных выступил открытое API LiveDune. Проведен анализ предметной области, технологии предобработки и формирования моделей, основные подходы к проектированию регрессионных моделей. Разработана математическая модель анализа, которая легла в основу искусственной нейронной сети.

Основными заинтересованными лицами системы скоринга являются рекламодатели и маркетинговые агентства. По результатам реализации проекта выявлена инновационность проекта, отсутствие на рынке прямых конкурентов.

Общие затраты на реализацию научно-исследовательской работы составляют 191 512,23 руб.

Результаты работы могут быть использованы в качестве основы для дальнейшего анализа комплаентности аудитории блогеров для принятия решений о коллаборации с ним.

В качестве возможной реализации проекта на коммерческом рынке рассмотрена возможность слияния с Mail ru Group и LiveDune.

Список использованных источников

1. Epsilon research on services [Электронный ресурс] URL: <https://www.epsilon.com/us/about-us/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 02.03.2021г.)
2. Дюк, В. А. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях / В. А. Дюк, А. В. Флегонтов, И. К. Фомина // 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:25.05.2021).
3. Aurelien Geron. Hands-On Machine Learning with Sklearn and Tensorflow.– O'Reilly, 2015: - 751p.
4. Чоккой, В. З. Обработка и разведочный анализ числовых массивов данных/ В. З Чоккой // Crede Experto: транспорт, общество, образование, язык. 2017. №3. URL: <https://cyberleninka.ru/article/n/obrabotka-i-razvedochnyu-analiz-chislovyh-massivov-dannyh> (дата обращения: 21.03.2021).
5. Решетникова Н. В. Особенности использования предобработки данных для изучения динамики их изменения с помощью нейросетей с учителем // Новые информационные технологии в автоматизированных системах. / Н. В. Решетникова, Л. А. Жуков 2005. №8. URL: <https://cyberleninka.ru/article/n/osobennosti-ispolzovaniya-predobrabotki-dannyh-dlya-izucheniya-dinamiki-ih-izmeneniya-s-pomoschyu-neyrosetey-s-uchitelem> (дата обращения:15.03.2021).
6. Bharath Ramsundar. Tensorflow for Deep Learning. – O'Reilly, 2015: - 253p.
7. Хашковский В. В. Современные подходы в организации систем обработки больших объемов данных / В. В. Хашковский, А. Н. Шкурко // Известия ЮФУ. Технические науки. 2014. №8. URL:

<https://cyberleninka.ru/article/n/sovremennye-podhody-v-organizatsii-sistem-obrabotki-bolshih-obemov-dannyh> (дата обращения: 17.03.2021).

8. Агафонов, А. А. Оценка и прогнозирование параметров транспортных потоков с использованием композиции методов машинного обучения и моделей прогнозирования временных рядов / А. А. Агафонов, В. В. Мясников // КО. 2014. №3. URL: https://books.google.ru/books?hl=en&lr&id=sm39DQAAQBAJ&oi=fnd&pg=PA11&dq=машинное%20обучение&ots=lmOKvoxKbA&sig=syblyAoh40RP8zbP-dTCZMkAmk&redir_esc=y%23v%3Donepage&q=машинное%20обучение&f=false (дата обращения: 25.05.2021).

9. Базенков, Н. И. Обзор информационных систем анализа социальных сетей / Н. И. Базенков, Д. А. Губанов // УБС. 2013. №41. URL: <https://cyberleninka.ru/article/n/obzor-informatsionnyh-sistem-analiza-sotsialnyh-setey> (дата обращения: 01.04.2021).

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, С. И. Интеллектуальный анализ данных для задач CRM // International Journal of Open Information Technologies. 2015. №2. URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-dlya-zadach-crm> (дата обращения: 04.04.2021).

14. Филатова, Т. В. 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: 04.04.2021).

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: 04.04.2021).

16. François Chollet. Deep Learning with Python. – Packt, 2011: - 547p.

17. Архангельская Е. В. «Глубокое обучение. Погружение в мир нейронных сетей», Библиотека программиста. / Е. В. Архангельская, Кадурин А., Николенко С. И. – 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. Игнатьев Н. А. О синтезе факторов в искусственных нейронных сетях//ЖВТ.2005.№3. URL: <https://cyberleninka.ru/article/n/o-sinteze-faktorov-v-iskusstvennyh-neuronnyh-setyah> (дата обращения: 09.04.2021).

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, П. Н. Программная реализация алгоритма градиентного бустинга деревьев решений / П. Н. Дружков, Н. Ю. Золотых, А. Н. Половинкин // Вестник ННГУ. 2011. №1.

URL:<https://cyberleninka.ru/article/n/programmnaayarealizatsiya-algoritma-gradientnogo-bustinga-dereviev-resheniy> (дата обращения: 01.05.2021).

25. Никулин, В. Н. Об однородных ансамблях при использовании метода бустинга в приложении к классификации несбалансированных данных/ В. Н. Никулин., С. А. Палешева // 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: 01.05.2021).

26. Охрана труда [Электронный ресурс] / Безопасность жизнедеятельности. URL: <http://www.grandars.ru/shkola/bezopasnost-zhiznedeyatelnosti/ohrana-truda.html>, свободный. Яз. Рус. (дата обращения: 16.05.2021.)

27. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы. Яз. Рус. (дата обращения: 16.05.2021.)

28. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений. Яз. Рус. (дата обращения: 16.05.2021.)

29. ГОСТ 12.1.038–82 Система стандартов безопасности труда. Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов. Яз. Рус. (дата обращения: 16.05.2021.)

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/, свободный. Яз. Рус. (дата обращения: 16.05.2021.)

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: 16.05.2021.)

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: 16.05.2021.)

33. ГОСТ Р 55090-2012 Ресурсосбережение. Обращение с отходами. Рекомендации по утилизации отходов бумаги. Яз. Рус. (дата обращения: 16.05.2021.)

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: 16.05.2021.)

35. Трудовой кодекс Российской Федерации от 30.12.2001 №197-ФЗ (ред. от 27.12.2018). – М.: Госстандарт России, 2018. – 176 с

Приложение А

(обязательное)

Матрица SWOT – анализа

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

<p>В1. Тенденция роста спроса на программный продукт. В2. Рост потребностей клиентов. В3. Захват смежных сегментов целевого рынка.</p>	<p>1. В1В2С1С3 – разработка различных версий программного продукта, ориентированных на более узкую целевую аудиторию. 2. В2В3С1С2С5 – Расширение списка анализируемых параметров. 3. В1В2В3С1С4 – реклама и продвижение продукта на смежном рынке.</p>	<p>1. В1В2Сл2Сл4 – расширение команды работников. 2. В3Сл3Сл5 – оптимизация продукта путем реализации гибкого интерфейса.</p>
<p>Угрозы: У1. Появление и развитие аналогичных систем. У2. Непопулярность продукта на рынке. У3. Сбои работы в различных средах функционирования продукта.</p>	<p>Угрозы развития: 1. У1С3С4С5 – использование широкого функционала и удобного интерфейса. 2. У2С1С2 – реклама и продвижение на целевом рынке</p>	<p>Уязвимости: 1. У3Сл1 – необходима дополнительная отладка для решения проблемы совместимости. 2. У1Сл4 – слияние с конкурирующим проектом</p>

Приложение Б

(обязательное)

Степень компетенций в реализации проекта

№ 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
	Итого баллов	49	48

Приложение В

(обязательное)

Вредные и опасные факторы, возникающие при работе с ПЭВМ

Факторы (ГОСТ 12.0.003– 2015)	Этапы работ			Нормативные документы
	Разработка	Изготовление	Эксплуатация	
1. Отклонение показателей микроклимата	+	+	+	СанПиН 2.2.4.548–96
2. Недостаточная освещенность рабочей зоны	+	+	+	СП 52.13330.2016
3. Повышенный уровень шума на рабочем месте	+	+	+	СН 2.2.4/2.1.8.562–96
4. Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	+	+	+	ГОСТ 12.1.038–82 ССБТ