

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (ТГУ)  
Институт прикладной математики и компьютерных наук

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК

Руководитель ОПОП

д-р техн. наук, профессор

 А.В. Замятин

« 03 » июня 2023 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА  
(МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

МОДЕЛЬ РАСПОЗНАВАНИЯ ВРЕДОНОСНОЙ ИНФОРМАЦИИ НА ОСНОВЕ  
КРАУДСОРСИНГА ПОЛЬЗОВАТЕЛЕЙ

по направлению подготовки 01.04.02 Прикладная математика и информатика,  
направленность (профиль) «Интеллектуальный анализ больших данных»

Чжу Сяофэн


Руководитель ВКР

д-р техн. наук, профессор

 А.В. Замятин

« 03 » июня 2023 г.

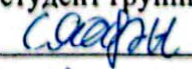
Консультант ВКР

 С.В. Карев

« 22 » мая 2023 г.

Автор работы

студент группы № 932128

 Сяофэн Чжу

« 03 » июня 2023 г.

Томск – 2023

Министерство науки и высшего образования Российской Федерации.  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)  
Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ  
Руководитель ОПОП  
д-р техн. наук, профессор

 А.В. Замятин  
подпись

« 10 » ноября 20 22 г.

ЗАДАНИЕ

по выполнению выпускной квалификационной работы магистра обучающегося  
Чжу Сяофэн

*Фамилия Имя Отчество обучающегося*

по направлению подготовки 01.04.02 Прикладная математика и информатика,  
направленность (профиль) «Интеллектуальный анализ больших данных»

1 Тема выпускной квалификационной работы

Модель распознавания вредоносной информации на основе краудсорсинга  
пользователей

2 Срок сдачи обучающимся выполненной выпускной квалификационной работы:

а) в учебный офис / деканат – 28.05.2023 б) в ГЭК – 07.06.2023

3 Исходные данные к работе:

Объект исследования – Вредоносная информация в социальных сетях.

Предмет исследования – Использование пользовательского краудсорсинга для  
распознавания вредоносной информации.

Цель исследования – Разработка модели распознавания вредоносной информации на  
основе пользовательского краудсорсинга.

Задачи:

1. Исследовать взаимосвязь между платформами социальных сетей и вредоносной  
информацией.

2. Изучить распространенные методы распознавания вредоносной информации.

3. Исследовать преимущества и недостатки общих методов распознавания.

4. Создать модель распознавания вредоносной информации на основе краудсорсинга  
пользователей.

5. Выбрать инструменты для создания симуляционных наборов данных.

6. Выбрать инструменты для создания симуляционных моделей.

7. Использовать симуляционных моделей и симуляционных наборов данных для  
проверки эффективности теоретической модели.

Методы исследования:

Поиск и анализ соответствующей литературы

Создание модели




## Тестирование симуляции

Организация или отрасль, по тематике которой выполняется работа, –  
Национальный исследовательский Томский государственный университет


### 4 Краткое содержание работы

1. Исследовались социальные сетевые платформы и вредоносная информация.
2. Изучались текущие методы распознавания вредоносной информации, вместе с их преимуществами и недостатками.
3. Создавались модель распознавания вредоносной информации на основе краудсорсинга пользователей.
4. Разрабатывались симуляционные наборы данных с помощью Python.
5. Создавались симуляционные модели с использованием Python.
6. Подтверждалась эффективность и точность модели с использованием симуляционных наборов данных и симуляционных моделей.

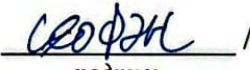
Научный руководитель выпускной  
квалификационной работы  
д-р техн. наук, директор ИПМКН НИ ТГУ  
*должность, место работы*

  
подпись / А.В. Замятин  
И.О. Фамилия

Консультант выпускной  
квалификационной работы  
ассистент кафедры ТОИ НИ ТГУ  
*должность, место работы*

  
подпись / С.В. Карев  
И.О. Фамилия

Задание принял к исполнению  
студент группы № 932128  
*должность, место работы*

  
подпись / Сяофэн Чжу  
И.О. Фамилия

## АННОТАЦИЯ

Данная магистерская диссертация насчитывает 75 страниц, включает 10 рисунков и 11 таблиц, общее количество источников - 18.

Ключевые слова: СОЦИАЛЬНЫЕ СЕТИ, ВРЕДОНОСНАЯ ИНФОРМАЦИЯ, КРАУДСОРСИНГ ПОЛЬЗОВАТЕЛЕЙ, СИМУЛЯЦИОННАЯ МОДЕЛЬ.

Объектом исследования является вредоносная информация в социальных сетях.

Предмет исследования является использование краудсорсинга пользователей для выявления вредоносной информации.

Целью исследования является создание модели распознавания вредоносной информации на основе краудсорсинга пользователей и проверка ее эффективности с помощью симуляционных наборов данных и симуляционных моделей.

Для достижения поставленных целей необходимо решить следующие задачи:

1. Исследовать взаимосвязь между платформами социальных сетей и вредоносной информацией.
2. Изучить распространенные методы распознавания вредоносной информации.
3. Исследовать преимущества и недостатки общих методов распознавания.
4. Создать модель распознавания вредоносной информации на основе краудсорсинга пользователей.
5. Выбрать инструменты для создания симуляционных наборов данных.
6. Выбрать инструменты для создания симуляционных моделей.
7. Использовать симуляционных моделей и симуляционных наборов данных для проверки эффективности теоретической модели.

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	3
1 Социальные сети и вредоносная информация .....	5
1.1 Платформы социальных сетей .....	5
1.2 Вредоносная информация в социальных сетях .....	7
1.3 Распространенные методы распознавания вредоносной информации .....	11
2 Модель распознавания вредоносной информации на основе краудсорсинга пользователей .....	18
2.1 Обзор Модели .....	18
2.2 Первая часть модели: Механизм оценки на уровне пользователя .....	23
2.3 Вторая часть модели: Механизм отбора краудсорсинга .....	30
2.4 Третья часть модели: Механизм расчета результатов краудсорсинга .....	33
2.5 Четвертая часть модели: Механизм обработки споров в краудсорсинге .....	37
2.6 Пятая часть модели: Механизм обработки сообщенного контента .....	42
2.7 Шестая часть модели: Механизм расчета баллов краудсорсинга .....	44
3 Валидация с помощью симуляции .....	48
3.1 Дизайн набора данных для симуляции .....	48
3.2 Дизайн модели симуляции .....	51
3.3 Анализ результатов симуляции .....	56
3.4 Сравнение моделей .....	59
ЗАКЛЮЧЕНИЕ .....	71
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ .....	73

## **ВВЕДЕНИЕ**

С появлением и быстрым развитием социальных сетей все больше пользователей наслаждаются удобством и преимуществами, которые они приносят. Однако развитие социальных платформ также привело к многим проблемам, среди которых особенно выделяется проблема вредоносной информации, которая стала крупным вызовом.

В ответ на вредоносную информацию, социальные сетевые платформы разработали различные методы распознавания, каждый из которых имеет свои достоинства и недостатки. Однако ни один единственный метод не может полностью сдержать распространение вредоносной информации.

Из-за огромного ущерба, наносимого вредоносной информацией, особенно с потенциалом вызвать вредные социальные события, точная распознавание вредоносной информации в социальных сетях имеет первостепенное значение.

Учитывая важность данного исследовательского поля, целью этой статьи является построение модели распознавания вредоносной информации на основе краудсорсинга пользователей и подтверждение ее эффективности с помощью симуляционных наборов данных и симуляционных моделей.

Задачи, связанные с достижением этой цели, следующие:

1. Исследовать взаимосвязь между платформами социальных сетей и вредоносной информацией.
2. Изучить распространенные методы распознавания вредоносной информации.
3. Исследовать преимущества и недостатки общих методов распознавания.
4. Создать модель распознавания вредоносной информации на основе краудсорсинга пользователей.
5. Выбрать инструменты для создания симуляционных наборов данных.

6. Выбрать инструменты для создания симуляционных моделей.

7. Использовать симуляционных моделей и симуляционных наборов данных для проверки эффективности теоретической модели.

Основное внимание в исследовании уделяется модели распознавания вредоносной информации на основе краудсорсинга пользователей для распознавания вредоносной информации на социальных сетевых платформах.

# **1 Социальные сети и вредоносная информация**

## **1.1 Платформы социальных сетей**

С прогрессом и развитием науки и технологии, приход эпохи интернета привнес множество изменений в жизнь человека. В эпоху интернета, появление платформ социальных сетей представляет собой значительное технологическое продвижение.

Платформы социальных сетей, также известные как социальные медиа, относятся к платформам производства и обмена контентом пользователей в интернете. Это инструменты и платформы, которые используют люди для обмена мнениями, взглядами, опытом и точками зрения[1].

С развитием интернета, платформы социальных сетей стали постепенно диверсифицироваться от их первоначальной однородной формы. В зависимости от различных услуг и функций, которые платформы социальных сетей предоставляют своим пользователям, они обычно классифицируются на пять различных категорий[2]:

1. Платформы распространения контента: Эти платформы в основном используются для распространения различных типов контента, включая текст, изображения, видео и т. д. Сюда входят такие платформы как Facebook, Twitter, Instagram, и TikTok. Эти платформы позволяют пользователям широко делиться и распространять контент, который они создают.

2. Платформы мгновенного обмена сообщениями: Эти платформы обеспечивают реальное общение между пользователями через текст, изображения и другие функции чата. Типичные примеры таких платформ - WhatsApp и WeChat. Эти платформы позволяют пользователям общаться в режиме реального времени без географических ограничений.

3. Платформы для создания длинных текстов: Эти платформы предоставляют пользователям цифровые пространства для создания,



публикации и управления статьями или блогами. Репрезентативные примеры включают WordPress и Blogger.

4. Платформы общественного обсуждения: Эти платформы позволяют пользователям генерировать темы для обсуждения и участвовать в общественных беседах. Репрезентативные примеры - Reddit и Quora.

5. Платформы комментирования контента: Эти платформы предоставляют пользователям платформу для создания и обмена оценками или комментариями на различные темы или объекты. Репрезентативные примеры включают IMDB и Rotten Tomatoes.

Прогресс и развитие платформ социальных сетей принесли много пользы людям, и эти преимущества постоянно влияют на различные аспекты жизни каждого человека[3].

Во-первых, с индивидуальной точки зрения, появление платформ социальных сетей, особенно платформ мгновенного обмена сообщениями, значительно усиливает взаимодействие и связность между людьми, повышая возможности коммуникации до новых вершин. Теперь люди со всего мира могут общаться друг с другом в любое время и в любом месте через интернет и платформы социальных сетей. Пространственное расстояние и географические границы больше не являются препятствием для человеческого общения.

Во-вторых, с бизнес-точки зрения, большая и разнообразная база пользователей платформ социальных сетей представляет большую ценность для коммерческих компаний. Все больше и больше предприятий склоняются к использованию платформ социальных сетей для своих деятельности, что создает основу для продвижения развития и инноваций в бизнесе.

Наконец, с общественной точки зрения, платформы социальных сетей предоставляют платформу для обмена информацией. Пользователи могут быть в курсе последних новостей и тенденций со всего мира через платформы социальных сетей. Кроме того, пользователи могут легко

получить доступ к новым знаниям и информации или получить глубокое понимание перспектив, опыта и жизненных ситуаций других пользователей. Платформы социальных сетей, безусловно, способствуют глобальному обмену информацией и точками зрения.

Однако, вместе с удобствами и преимуществами, которые приносят платформы социальных сетей в жизнь людей, они также порождают целый ряд проблем[4].

Например, современные платформы социальных сетей переполнены большим количеством вредоносной информации. Эта вредоносная информация не только портит пользовательский опыт на платформах социальных сетей, но и может привести к серии серьезных социальных проблем. Вредоносная информация была всепроникающей проблемой на платформах социальных сетей с момента их появления, представляя существенную угрозу для их надежности. В результате все больше и больше исследователей сосредотачиваются на том, как определить вредоносную информацию в социальных сетях.

## **1.2 Вредоносная информация в социальных сетях**

Вредоносная информация в социальных сетях относится к информации, созданной пользователями с намерением нанести вред другим пользователям или самой платформе. Вредоносная информация может принимать форму злобных комментариев, клеветы, слухов, издевательств, нападок, преследований, угроз, разжигания ненависти или других форм недопустимого контента. Создатели и распространители вредоносной информации, как правило, стремятся обмануть, использовать или нанести вред другим пользователям[5].

На основании содержания вредоносной информации, ее можно разделить на следующие четыре типа[6 , 7]:

1. Вредоносная информация с оскорбительным содержанием: Этот тип вредоносной информации включает оскорбления, угрозы, запугивание, преследование и аналогичные формы. Как правило, создатели вредоносной информации с оскорбительным содержанием стремятся вербально атаковать других пользователей, нанося вред их психическому здоровью.

Вредоносная информация с оскорбительным содержанием - это самый распространенный тип вредоносной информации на платформах социальных сетей и обычно имеет умеренный уровень вреда.

Например, использование ненормативной лексики часто считается вредоносной информацией с оскорбительным содержанием. Обычно это рассматривается как вербальное насилие в отношении других пользователей. Однако его степень оскорбительности обычно минимальна для большинства пользователей, что приводит к минимальному вреду.

Однако некоторая вредоносная информация с оскорбительным содержанием может нанести значительный вред. Например, угрозы или сексуальные домогательства часто считаются вредоносной информацией с оскорбительным содержанием, но эти контенты могут нанести значительный физический и психический вред другим. Более того, в некоторых странах отправка контента, содержащего угрозы или сексуальные домогательства, считается нарушением закона.

2. Вредоносная информация с ложной информацией: Это слухи, мошенничества, ложь и т. д., которые могут ввести в заблуждение других пользователей, наткнувшихся на эту информацию, повлияв на их суждение. Вред, причиненный вредоносной информацией с ложной информацией, зависит от скорости и масштаба ее распространения. Когда вредоносная информация с ложной информацией распространяется быстро и широко, это часто приводит к злободневным социальным событиям, вызывая значительный вред.

Например, в 2011 году, во время утечки в атомной электростанции Фукусима в Японии, кто-то сплел слух на китайской платформе социальных сетей Weibo, утверждая, что морская вода у берегов Китая была загрязнена ядерной радиацией и что поваренная соль станет дефицитным товаром. Этот слух быстро распространился в большом масштабе, что привело к паническим покупкам соли, взлету цен на соль и, до некоторой степени, социальной панике. В этом вредоносном социальном событии некоторые люди даже купили 13 000 фунтов соли по завышенным ценам.

3. Вредоносная информация с разжигающим ненависть содержанием: Этот тип информации может разжигать ненависть, способствовать социальному разделению, нападать или подстрекать других нападать на определенную группу или отдельного человека. Вредоносная информация с разжигающим ненависть содержанием часто имеет высокий уровень вреда.

В отличие от вредоносной информации с оскорбительным содержанием, издатели вредоносной информации с разжигающим ненависть содержанием стремятся подстрекать или провоцировать больше людей присоединиться к атаке на определенную группу или отдельного человека. Вредоносная информация с разжигающим ненависть содержанием часто включает дискриминацию определенной группы или отдельного человека с намерением уничтожить эту группу или человека.

Например, белые расисты в Соединенных Штатах часто публикуют вредоносную информацию с разжигающим ненависть содержанием, нацеленную на людей цвета.

Вредоносная информация с разжигающим ненависть содержанием часто имеет сильный уровень агрессии и может нанести значительный вред жертвам или целевым группам.

4. Мошенническая вредоносная информация: Этот тип информации разработан для обманных целей и часто включает ложную информацию или обещания, чтобы заманить других пользователей в ловушку. Мошенническая

вредоносная информация охватывает широкий спектр контента, от поддельных сообщений о лотереях до вводящих в заблуждение бизнес-предложений.

В отличие от вредоносной информации с ложной информацией, мошенническая вредоносная информация имеет специфическую направленность, с целью обмануть конкретных пользователей или группы на деньги.

Вред, причиненный мошеннической вредоносной информацией, чрезвычайно серьезен, потому что его целью является обман жертв с целью получения платежей или предоставления личной информации. Как только жертвы попадают на этот тип информации, они обычно страдают значительными убытками[8].

Например, одной из распространенных форм мошеннической информации является подделка банка или другой авторитетной организации, когда отправитель просит получателя подтвердить данные своего счета, тем самым получая доступ к его конфиденциальной информации.

Вред от вредоносной информации в социальных сетях неоспорим. Будь то оскорбительный контент, ложная информация или речь, разжигающая ненависть, эти вредоносные информации могут оказать глубокое влияние на отдельных лиц, включая психологический ущерб и различные проблемы в их реальной жизни, такие как введение в заблуждение ложной информацией и экономические потери.

Для общества в целом вредоносная информация также имеет значительные негативные последствия. Распространение ложной информации и речи, разжигающей ненависть, может разжигать конфликты и дестабилизировать общество.



### **1.3 Распространенные методы распознавания вредоносной информации**

С развитием социальных сетей, вред от вредоносной информации для общества становится все более серьезным. Таким образом, одной из важных целей платформ социальных сетей является разработка методов, которые могут быстро и точно распознавать вредоносную информацию. Для достижения этой цели, исследователи и технические специалисты предложили множество различных методов распознавания вредоносной информации. Вот некоторые из наиболее распространенных методов распознавания вредоносной информации:

1. Фильтрация на основе ключевых слов[9]: Фильтрация на основе ключевых слов - это базовый подход к определению вредоносной информации. Его основной принцип заключается в фильтрации содержимого на основе заранее определенного набора оскорбительных, агрессивных или чувствительных слов и фраз. Когда контент, который пользователь создал, содержит эти ключевые слова, платформа автоматически помечает или обрабатывает информацию.

Фильтрация на основе ключевых слов - это легко понятный и простой метод для распознавания прямой вредоносной информации. Например, для вредоносной информации с оскорбительным содержанием платформа обычно фильтрует слова или предложения с оскорбительным языком. Если эти слова или предложения обнаруживаются в контенте, созданном или распространяемом пользователем, информация немедленно помечается или обрабатывается автоматически.

Однако этот метод имеет значительные ограничения. Во-первых, он не может идентифицировать вредоносную информацию, которая не содержит конкретных ключевых слов, но все же носит вредоносный характер, такой как тонкие нападки или ирония. Во-вторых, он не решает семантических проблем

в вредоносной информации, таких как контекст или двусмысленность. Наконец, фильтрация по ключевым словам может привести к чрезмерной цензуре, когда информация, содержащая ключевые слова, но без фактического вредоносного намерения, может быть ошибочно отфильтрована.

Несмотря на свои ограничения, фильтрация на основе ключевых слов все еще широко используется многими платформами социальных сетей из-за ее простоты внедрения.

2. Метод ручной модерации[10]: использование ручной модерации для распознавания, является ли информация вредоносной, сейчас является методом с самой высокой точностью распознавания. Особенно, модераторы, прошедшие специальное обучение, обладают очень высокой точностью распознавания.

Однако этот метод связан с огромными трудовыми затратами, особенно для социальных платформ с большим числом пользователей, которые требуют нанимать большое количество модераторов для проверки, что приводит к высоким операционным затратам. Более того, процесс ручной модерации также столкнулся с проблемами субъективности и предрассудков, так как разные модераторы могут иметь разные критерии для распознавания вредоносной информации. В то же время, этот метод также оставляет полное толкование вредоносной информации на усмотрение социальной платформы и нанятых ими модераторов. Некоторые пользователи могут считать, что некоторая информация является вредоносной, но социальная платформа и нанятые ими модераторы могут не считать ее вредоносной. В этом случае, является ли информация вредоносной, полностью зависит от социальной платформы и ее модераторов.

Ручная модерация является методом, который сейчас используется на каждой платформе социальных сетей, но обычно он используется в сочетании с другими методами распознавания, потому что стоимость распознавания при ручной модерации очень высока.

3. Методы машинного и глубокого обучения[11]: Использование методов машинного и глубокого обучения для распознавания - более продвинутый подход к обнаружению вредоносной информации. Этот метод основан на обучении моделей с помощью большого количества размеченных данных, позволяя моделям учиться на основе признаков вредоносной информации и делать суждения на их основе. Методы машинного и глубокого обучения в настоящее время являются наиболее широко используемыми стратегиями для распознавания вредоносной информации.

Распространенные методы машинного обучения, используемые для распознавания вредоносной информации, включают классификаторы Наивного Байеса, метод опорных векторов, деревья решений, и методы глубокого обучения, такие как сверточные нейронные сети и рекуррентные нейронные сети.

Преимущество использования методов машинного и глубокого обучения для распознавания в том, что они могут самостоятельно изучать признаки вредоносной информации на основе данных, и они могут обрабатывать более сложные семантические и контекстуальные проблемы. Кроме того, использование методов машинного и глубокого обучения для распознавания быстро и более точно по сравнению с фильтрацией по ключевым словам. Более того, использование этого подхода не вызывает проблем конфиденциальности для пользователей.

Однако недостатком этого подхода является то, что качество и количество обучающих данных значительно влияют на эффективность модели. Если обучающие данные низкого качества или недостаточного количества, точность обученной модели может быть низкой. Создание большого количества качественных обучающих данных требует значительного времени и ресурсов. Обучение моделей с использованием методов машинного и глубокого обучения также требует значительных вычислительных ресурсов.

На многих популярных платформах социальных сетей, использование технологий машинного обучения и глубокого обучения для распознавания вредоносной информации является очень популярным методом. Причина выбора этого метода заключается в том, что по сравнению с низкой эффективностью ручного распознавания, технологии машинного обучения и глубокого обучения могут немного уступать в точности, но их скорость распознавания очень высока. Особенно с развитием технологий машинного обучения и глубокого обучения, их точность распознавания постепенно увеличивается. В распознавании некоторых определенных категорий вредоносной информации, они могут приближаться к точности ручного распознавания.

4. Метод управления сообществом[12]: некоторые платформы социальных сетей, обычно общественные платформы, выбирают определенное количество членов сообщества и предоставляют им статус администраторов, позволяя этим пользователям помочь распознавать и обрабатывать вредоносную информацию внутри сообщества. Этот метод называется управлением сообществом.

Преимущество этого метода в том, что администраторы являются членами сообщества, они имеют глубокое понимание культуры и правил сообщества, что позволяет им более точно распознавать вредоносную информацию. Однако недостатки этого метода также очевидны. Во-первых, администраторы не прошли профессионального обучения, поэтому их точность может быть относительно низкой. Во-вторых, у администраторов могут быть личные предвзятые мнения, что может привести к субъективным суждениям. Кроме того, этот метод подходит только для определенных типов платформ социальных сетей.

Упомянутые выше четыре метода являются основными методами, используемыми для распознавания вредоносной информации. Каждый метод имеет свои преимущества и недостатки, и нет метода, который мог бы

полностью решить проблему вредоносной информации. В отношении использования машин для распознавания, представительным методом является использование технологий машинного обучения и глубокого обучения. Точность этих методов обычно зависит от объективных факторов, таких как используемые модели и обучающие наборы данных. В то же время, на объективном уровне, точность распознавания машинами все еще не сравнима с человеком, но человеческое распознавание подвержено различным субъективным факторам, которые могут привести к ошибкам.

Для многих платформ социальных сетей, комбинация различных методов распознавания обычно является одним из лучших вариантов. При выборе метода распознавания, платформы социальных сетей обычно ограничены размером и типом платформы. Платформы социальных сетей меньшего размера, вероятно, предпочтут использовать метод ручной модерации, поскольку они могут лучше справиться с относительно небольшим объемом информации и более сосредоточены на точности. Для платформ социальных сетей большого масштаба, внедрение автоматизированных технологий и методов машинного обучения может повысить эффективность обработки, хотя точность может немного снизиться. В то же время, стоимость также является одним из рассматриваемых факторов, для платформ социальных сетей меньшего размера, стоимость аппаратных и программных ресурсов, необходимых для обучения модели, часто велика, гораздо больше, чем стоимость ручной модерации. Для платформ социальных сетей большого масштаба, стоимость ручной модерации значительно превышает стоимость обучения модели.

С точки зрения типа платформы, разные платформы социальных сетей могут столкнуться с различными типами вредоносной информации[13]. Например, пользователи некоторых платформ могут быть более подвержены домогательствам вредоносной информации, содержащей ложную информацию, в то время как другие платформы могут столкнуться с большим



количеством мошеннической информации, направленной на пользователей. Поэтому при выборе метода распознавания, платформы социальных сетей должны учитывать свои конкретные риски, и разрабатывать наиболее подходящую стратегию распознавания на основе этих факторов.

Несмотря на непрерывное развитие технологий распознавания, распространение вредоносной информации не прекращается. Это в первую очередь связано с тем, что определение вредоносной информации - это чрезвычайно сложная проблема, которая выходит за рамки технических проблем и включает в себя социальные проблемы.

На техническом уровне распознавание вредоносной информации остается значительной проблемой. Несмотря на то, что современные технологии искусственного интеллекта и машинного обучения стали достаточно мощными, у них все еще есть ограничения в обработке текста и понимании семантики. Эти системы могут точно определить и отфильтровать некоторые прямые и явные формы вредоносной информации, такие как угрозы или оскорбления. Однако их точность уменьшается, когда речь идет об неявных, ироничных или тонких формах вредоносной информации. Кроме того, язык и выражения постоянно эволюционируют, и создатели вредоносной информации могут использовать новую лексику и способы выражения, чтобы избежать контроля на социальных сетевых платформах.

На социальном уровне социальные сети являются отражением реального общества и культуры, и создание вредоносной информации часто тесно связано с социальными и культурными проблемами в реальности. Например, если в сообществе есть серьезные проблемы дискриминации или предвзятости, эти проблемы могут проявиться в виде вредоносной информации на социальных сетевых платформах. В таких случаях полагаться только на технологические средства может оказаться неэффективным[14].

Кроме того, социальные сетевые платформы могут терпеть или даже способствовать распространению определенной вредоносной информации

из-за своих собственных интересов. Модели функционирования социальных сетевых платформ во многом зависят от участия и вовлеченности пользователей для получения рекламной выручки. Эта модель может создать серые зоны в распознавание вредоносной информации[15]. Например, для некоторых социальных сетей спорные и экстремальные материалы часто способствуют повышению взаимодействия и обмена между пользователями, тем самым увеличивая уровень участия пользователей и известность платформы. Эти платформы могут быть склонны терпеть такой контент, несмотря на риск присутствия вредоносной информации.

## **2 Модель распознавания вредоносной информации на основе краудсорсинга пользователей**

### **2.1 Обзор Модели**

Ручная модерация и управление сообществом - это два метода, которые зависят от человеческого вмешательства для распознавания вредоносной информации. Преимущество человеческого фактора заключается в его высокой объективной точности, однако он может подвергаться влиянию субъективных факторов, что может привести к снижению точности[16].

Кроме того, затраты на ручное распознавание всегда были достаточно высокими, что делает его подходящим только для использования в малом масштабе.

С другой стороны, в отношении скорости распознавания, если вредоносная информация появляется в больших количествах одновременно, скорость ручного распознавания может оказаться ограниченной, что приведет к низкой эффективности обработки вредоносной информации.

Если можно решить проблемы субъективного влияния, стоимости распознавания и скорости распознавания, использование ручной модерации для распознавания вредоносной информации станет идеальным выбором.

Использование краудсорсинга может одним из решений.

Краудсорсинг - это подход к принятию решений, основанный на коллективном интеллекте и консенсусе внутри группы[17]. Базовая идея краудсорсинга заключается в делегировании принятия решений группе, состоящей из многих членов, а не на определенном индивидууме или нескольких индивидуумах. Он основан на базовом предположении, что коллективные суждения обычно более точны и справедливы, чем суждения отдельных существ. Модель распознавания вредоносной информации на основе краудсорсинга пользователей применяет принципы краудсорсинга для распознавания вредоносной информации.

В отношении метода краудсорсинга, голосование многих пользователей может в значительной степени избежать ошибок, вызванных субъективностью отдельных лиц. В отношении скорости распознавания, хотя распознавание каждой вредоносной информации может занять больше времени, поскольку неизвестно, когда пользователь примет участие в краудсорсинге, однако когда нужно одновременно обрабатывать большое количество вредоносной информации, большое количество пользователей, участвующих в распознавании, может эффективно обрабатывать каждую вредоносную информацию параллельно, тем самым увеличивая скорость распознавания.

В отношении затрат на распознавание, участие пользователей в краудсорсинге обычно не требует от социальных сетевых платформ больших денежных затрат. Однако несомненно, что краудсорсинг требует больше трудозатрат, хотя эти трудовые ресурсы в основном бесплатны.

В социальных сетевых платформах метод краудсорсинга уже применяется в некоторых предшествующих случаях. Например, китайская социальная платформа Zhihu и платформа для отзывов Meituan уже внедрили систему краудсорсинга [18]. На платформе Zhihu определенные категории жалоб на контент были определены как объекты для краудсорсинга. Платформа выбрала хорошо работающих пользователей для участия в процессе краудсорсинга. Жалобы на контент были случайным образом распределены среди выбранных пользователей для распознавания.

В процессе краудсорсинга пользователи голосуют. Возможные варианты для пользователя разделены на две категории: согласен, что это вредоносная информация, и не согласен, что это вредоносная информация. Для каждого задания краудсорсинга, как только количество действительных голосов достигает 30, система вычисляет результат краудсорсинга на основе голосов пользователей. Если количество голосов за один вариант превышает 60% общего количества действительных голосов, этот вариант считается

результатом краудсорсинга. Если ни один вариант не получает больше 60% голосов, задание краудсорсинга считается проваленным и переходит на следующий раунд краудсорсинга.

Модель краудсорсинга Zhihu полностью зависит от результатов голосования пользователей и количества голосов для расчета результата краудсорсинга, не принимая во внимание другие факторы, такие как надежность пользователя.

На основе модели краудсорсинга Zhihu, это исследование предлагает модель распознавания вредоносной информации на основе краудсорсинга. В этой модели пользователи могут жаловаться на определенный контент как на возможную вредоносную информацию или определять определенную информацию как возможную вредоносную информацию другими способами. Возможная вредоносная информация передается выбранной группе пользователей для просмотра в качестве задания краудсорсинга, эти пользователи выбираются моделью и называются краудсорсеры. Эти краудсорсеры голосуют за то, считать ли информацию вредоносной. В конце концов, результаты голосования краудсорсеров будут привесованы и рассчитаны, а также будет рассчитана степень доверия к результатам. Затем в соответствии с уровнем доверия и результатами голосования проводится соответствующая обработка.

Специфические механизмы модели распознавания вредоносного контента на основе краудсорсинга пользователей следующие:

Во-первых, модель оценивает и фильтрует пользователей, чтобы определить тех, кто имеет право сообщать и участвовать в краудсорсинге. Эта часть модели называется механизмом оценки пользователей.

Затем модель обрабатывает потенциально вредоносный контент, сообщенный пользователями или идентифицированный другими способами. Для такого потенциально вредоносного контента модель выбирает пользователей для участия в процессе краудсорсинга на основе различных



факторов. Эта часть модели называется механизмом выбора для краудсорсинга.

Затем модель рассчитывает вес каждого голоса пользователя на основе голосов, поданных участвующими пользователями, в результате чего получается предварительный результат краудсорсинга. Эта часть модели называется механизмом расчета результатов краудсорсинга.

После этого модель оценивает надежность предварительного результата краудсорсинга. Если модель определяет высокую надежность, она отмечает предварительный результат как окончательный результат краудсорсинга и предпринимает соответствующие действия. Одновременно модель учитывает несколько факторов, чтобы определить, могут ли пользователи возражать против окончательного результата краудсорсинга. Если модель считает, что надежность предварительного результата низка и недостаточна для действия, она переходит к следующему раунду краудсорсинга. Эта часть модели называется механизмом разрешения споров по краудсорсингу, а обработка сообщенного контента называется механизмом обработки сообщенного контента.

Наконец, после обработки вредоносного контента, модель рассчитывает баллы краудсорсинга для участвующих пользователей в качестве наград или штрафов. Эта часть модели называется механизмом расчета баллов краудсорсинга.

Рисунок 1 наглядно демонстрирует общую структуру модели:

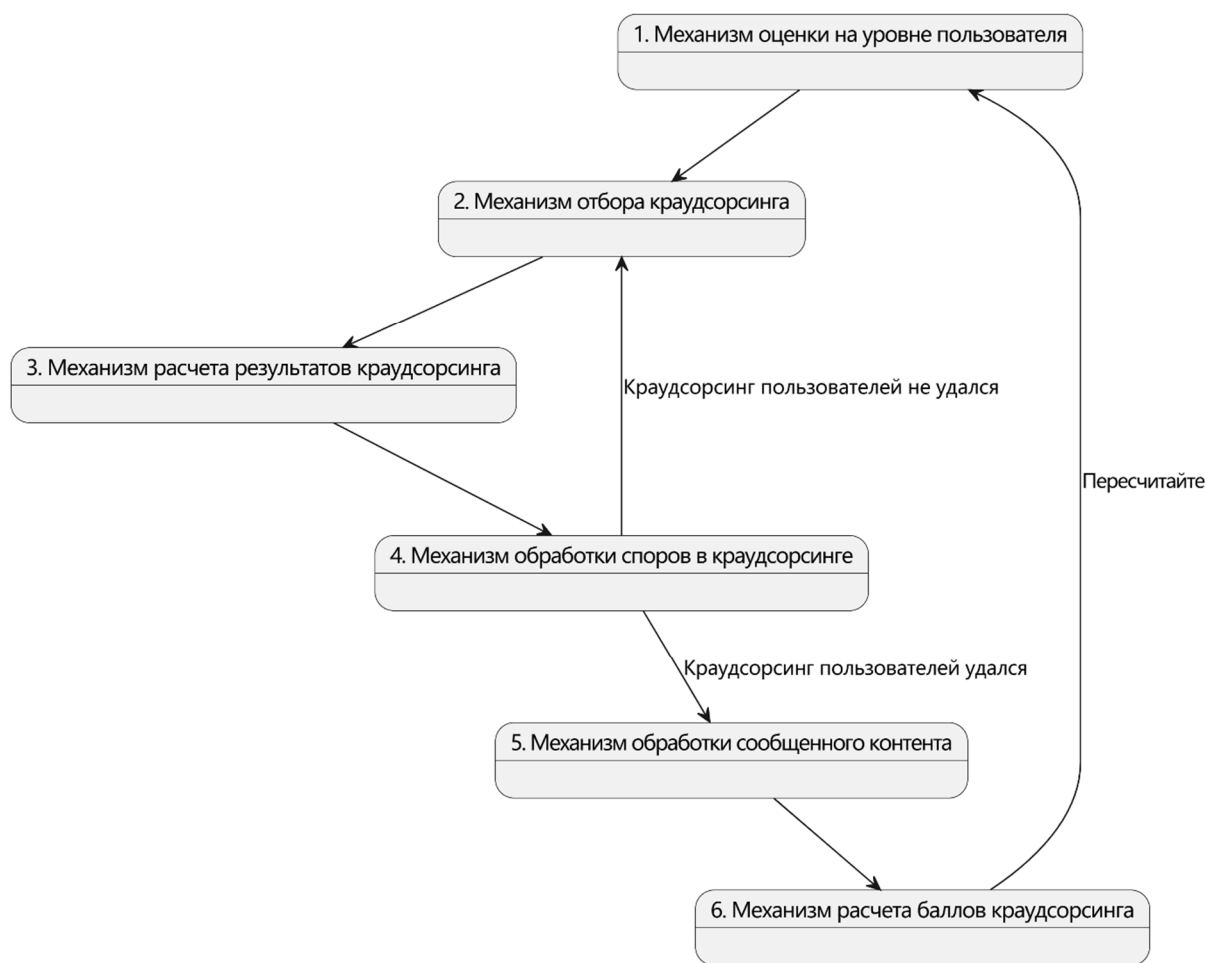


Рисунок 1 – Обзор Модели

Ключевым преимуществом модели распознавания вредоносного контента на основе краудсорсинга пользователей является ее способность использовать коллективный интеллект, улучшая точность и справедливость распознавания вредоносного контента. Специфические преимущества и недостатки модели следующие:

Преимущества модели:

- Коллективный интеллект: Пользователи, участвующие в краудсорсинге, происходят из различных сред и имеют разные точки зрения, что позволяет модели извлекать пользу из различных взглядов при суждении о значении и намерениях информации, тем самым улучшая точность распознавания вредоносного контента.

- Справедливость: Решения на основе краудсорсинга, а не единого принятия решений, улучшают справедливость суждения.

- Взвешенный расчет: При расчете весов голосования модель присваивает больший вес пользователям с большей достоверностью, улучшая точность и надежность модели.

Недостатки модели:

- Медленная скорость распознавания: скорость распознавания вредоносного контента относительно медленная, потому что пользователи могут не быть немедленно доступны для участия в краудсорсинге.

- Потенциальные групповые предрассудки: Если большинство пользователей имеют определенные предрассудки, это может повлиять на результаты краудсорсинга и уменьшить справедливость модели.

В то же время модель сталкивается с потенциальными проблемами. Например, некоторые пользователи могут попытаться манипулировать результатами голосования. Чтобы предотвратить такие ситуации, модель случайным образом выбирает пользователей для краудсорсинга и гарантирует, что участвующие пользователи не знают друг о друге идентичности.

Модель также полагается на активное участие пользователей. Если пользователи не хотят участвовать или не обладают способностью судить о вредоносном контенте, точность модели уменьшится. Чтобы решить эту проблему, модель предоставляет дополнительные стимулы активно участвующим пользователям, чтобы привлечь больше пользователей к участию в краудсорсинге.

## **2.2 Первая часть модели: Механизм оценки на уровне пользователя**

Первая часть модели отвечает за оценку пользователей и присвоение им уровней пользователей. Уровень пользователя - один из основных

механизмов модели, рассчитанный на основе поведения пользователя и представленный в виде числа с плавающей запятой от 0 до 100.

На основе уровня пользователя пользователи разделены на следующие пять классов:

1. Уровень пользователя между от 0 до 20: Пользователи, которых модель определяет как вредоносных. Эти пользователи заблокированы и не могут создавать или делиться каким-либо контентом.

2. Уровень пользователя между от 20 до 50: Пользователи, которых модель определяет как потенциально вредоносных. Для потенциально вредоносных пользователей действуют следующие ограничения:

- Невозможно сообщить о контенте других пользователей, содержащем вредоносную информацию.
- Невозможно участвовать в краудсорсинге.
- Высокая вероятность одобрения контента краудсорсингом, если он сообщается как содержащий вредоносную информацию.
- Меньший шанс быть допущенным к подаче возражений, когда модель оценивает возможность подачи возражений.

3. Уровень пользователя между от 50 до 60: Пользователи, которых модель определяет как подозрительных. Для подозрительных пользователей действуют следующие ограничения:

- Невозможно сообщить о контенте других пользователей, содержащем вредоносную информацию.
- Невозможно участвовать в краудсорсинге.

4. Уровень пользователя между от 60 до 70: Пользователи, которых модель определяет как обычных. У обычных пользователей есть следующие права:

- Разрешено сообщать о контенте других пользователей, содержащем вредоносную информацию.

Однако, у них есть следующее ограничение:

- Невозможно участвовать в краудсорсинге.

5. Уровень пользователя между от 70 до 100: Пользователи, которых модель определяет как доверенных. У доверенных пользователей есть следующие права:

- Разрешено сообщать о контенте других пользователей, содержащем вредоносную информацию.

- Разрешено участвовать в краудсорсинге.

- Высокая вероятность одобрения сообщенного контента краудсорсингом.

- Модель более осторожно судит, когда контент, созданный ими, сообщается как содержащий вредоносную информацию.

- Большой шанс на то, что им будет разрешено подать возражения на результаты краудсорсинга.

Результаты классификации пользователей по уровню представлены в следующей таблице:

Таблица 1 – Результаты классификации пользователей

Уровень пользователя	Описание	Можно ли сообщить	Можно ли участвовать в краудсорсинге
0-20	Пользователи-вредоносные	Нет	Нет
20-50	Потенциально вредоносные	Нет	Нет

Окончание таблицы 1

Уровень пользователя	Описание	Можно ли сообщить	Можно ли участвовать в краудсорсинге
50-60	Подозрительные пользователи	Нет	Нет
60-70	Обычные пользователи	Да	Нет
70-100	Доверенные пользователи	Да	Да

Конкретный механизм оценки уровня пользователя следующий:

1. Для новых пользователей модель назначает начальный уровень 50. Когда новый пользователь проходит аутентификацию личной информации, его уровень повышается до 60. Если новый пользователь не проходит аутентификацию личной информации, верхний предел их уровня устанавливается как 59. Конкретная диаграмма процесса представлена на Рисунке 2.

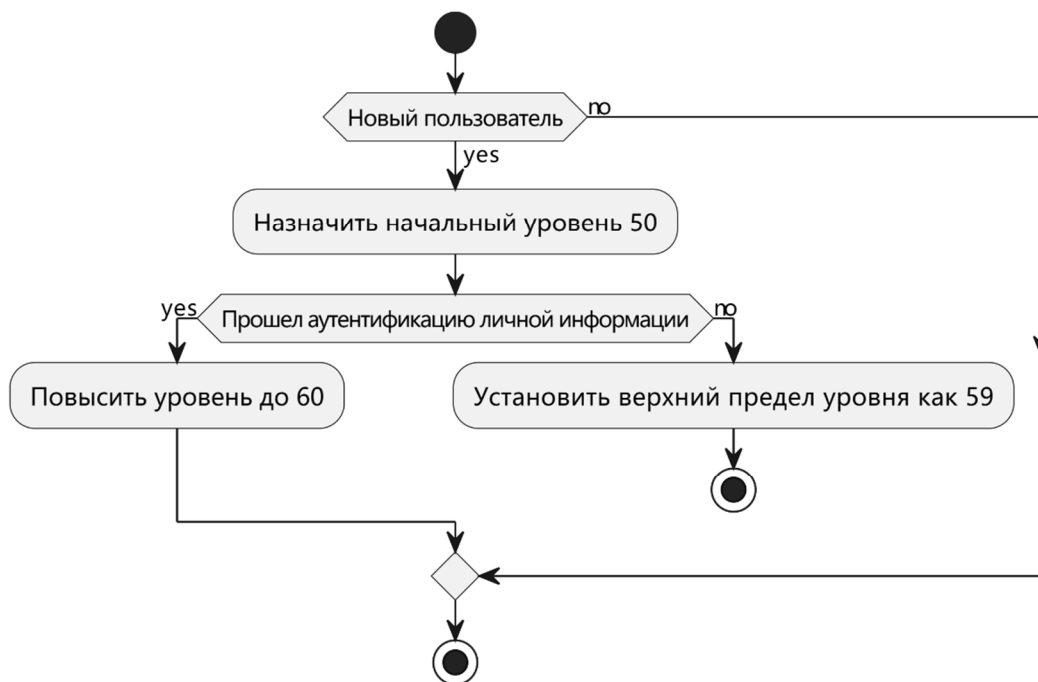


Рисунок 2 – Оценка новых пользователей

2. Пользователи зарабатывают баллы активности через активное поведение. Если пользователь признан активным пользователем в данный день, он получает баллы активности в виде числа с плавающей запятой в диапазоне от 0 до 1. Однако, если пользователь обнаружен создающим или распространяющим контент, содержащий вредоносную информацию в этот день, он не имеет права на баллы активности. Активные действия включают вход в социальную сеть, просмотр контента и т.д.

Значение баллов активности связано с уровнем пользователя:

- Пользователи с уровнем между 20 и 50 получают меньше баллов активности, и чем ниже их уровень, тем меньше баллов.
- Пользователи с уровнем между 50 и 70 получают больше баллов активности, и чем выше их уровень, тем больше баллов.
- Пользователи с уровнем между 70 и 100 получают меньше баллов активности, и чем выше их уровень, тем меньше баллов.

С помощью этой системы, пользователи с уровнем между 50 и 70 могут увеличить свой уровень через ежедневную активность. Для пользователей с уровнями выше 70 модель поощряет их улучшать свой уровень, участвуя в краудсорсинге, поэтому модель предоставляет им только меньшие баллы активности.

Конкретная диаграмма процесса представлена на Рисунке 3.



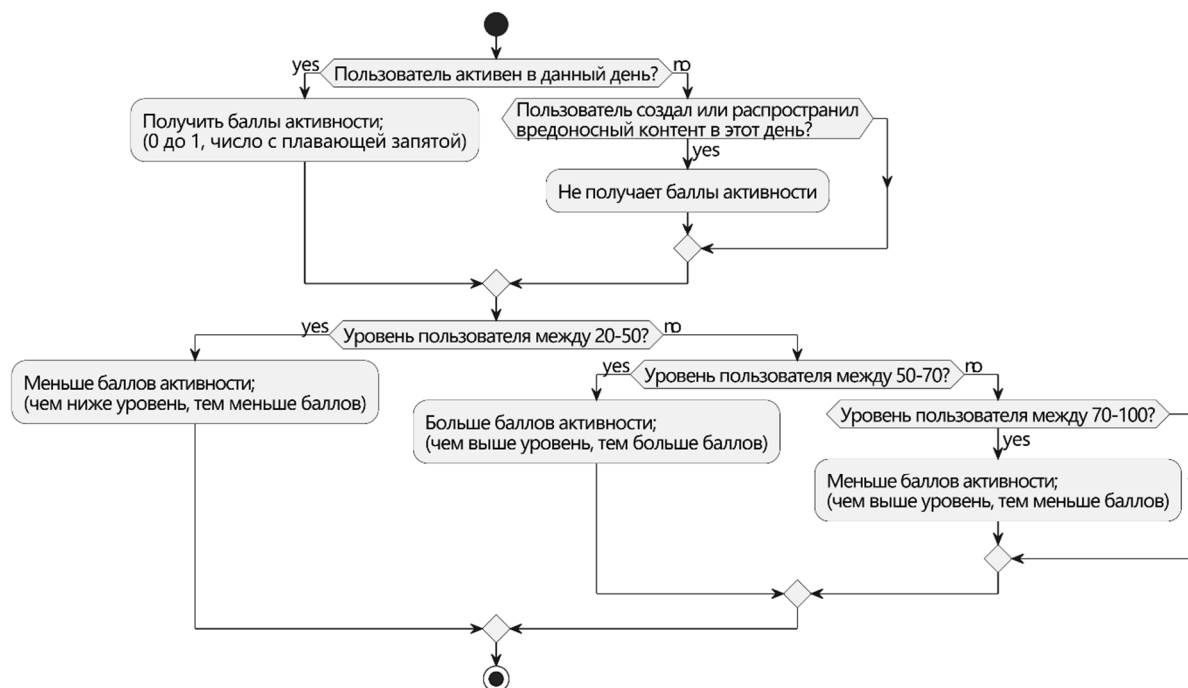


Рисунок 3 – Расчет активного балла пользователя.

3. Пользователи получают баллы за активное и ответственное сообщение о вредоносной информации. Если пользователь сообщает о контенте, содержащем вредоносную информацию, и контент оценивается краудсорсинговой группой, выбранной моделью, как действительно содержащий вредоносную информацию, пользователь получает положительный балл за сообщение. Однако, если сообщенный контент оценивается краудсорсинговой группой как не содержащий вредоносной информации и среди участников мало несогласия, сообщавший пользователь получает отрицательный балл за сообщение. Если пользователь сообщает о большом количестве контента за короткий период, и более двух третей контента оценивается краудсорсинговой группой как не содержащий вредоносной информации, уровень пользователя сбрасывается до 50.

Конкретная диаграмма процесса представлена на Рисунке 4.

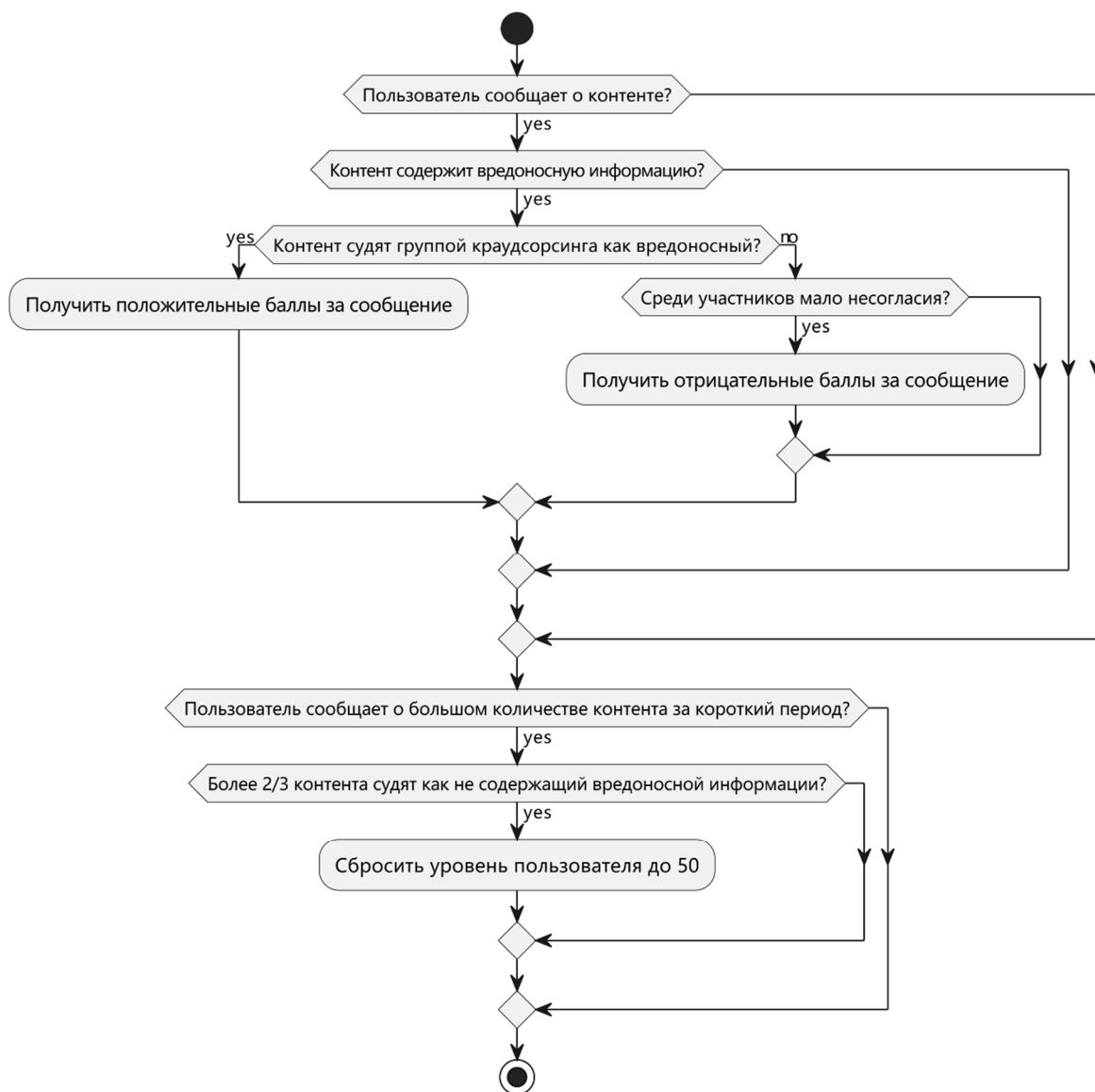


Рисунок 4 – Получение баллов за сообщение о вредоносной информации.

4. Пользователи получают штрафные баллы, когда контент, который они создали, судится группой краудсорсинга как содержащий вредоносную информацию. Если контент, который пользователь создал, сообщается и окончательный результат краудсорсинга определяет, что он содержит вредоносную информацию, пользователь получает штрафной балл, представленный в виде отрицательного значения. Конкретное значение штрафного балла рассчитывается механизмом модели обработки сообщенного контента.

5. Пользователи зарабатывают баллы краудсорсинга через участие в краудсорсинге. Когда уровень пользователя выше 70, они могут участвовать в краудсорсинге. Для задачи краудсорсинга, если голос пользователя совпадает с окончательным результатом краудсорсинга, они получают положительный балл краудсорсинга. Если голос пользователя отличается от окончательного результата краудсорсинга, они получают отрицательный балл краудсорсинга. Конкретное значение балла краудсорсинга рассчитывается механизмом наград и наказаний модели краудсорсинга.

### **2.3 Вторая часть модели: Механизм отбора краудсорсинга**

Вторая часть модели - это механизм отбора для краудсорсинга, который определяет, сколько пользователей должно участвовать в судействе при обработке сообщенного контента и какие пользователи должны быть выбраны для краудсорсинга.

При определении количества пользователей для участия в краудсорсинге модель учитывает следующие факторы:

1. Тип сообщаемого контента: когда пользователи сообщают о контенте, они обязаны указать тип вредоносной информации, к которому, по их мнению, относится этот контент. Затем модель использует исторические данные для распознавания базовой сложности распознавания каждого типа вредоносной информации. Для типов с более высокой сложностью модель выбирает больше пользователей для участия в краудсорсинге.

2. Уровень пользователя, о котором сообщили: чем выше уровень пользователя, о котором сообщили, тем больше пользователей модель выбирает для краудсорсинга. Это связано с тем, что пользователи с более высокими уровнями считаются надежными пользователями, и модель более осторожно обрабатывает сообщения, связанные с надежными пользователями. Выбор большего количества пользователей для участия в краудсорсинге снижает вес голоса каждого пользователя в процессе голосования

краудсорсинга, снижая влияние отдельных ошибочных результатов голосования на окончательный результат краудсорсинга.

3. Уровень пользователя, который сообщает: чем выше уровень пользователя, который сообщает, тем меньше пользователей модель выбирает для краудсорсинга. Это связано с тем, что пользователи с более высокими уровнями считаются надежными пользователями, и их суждение о вредоносной информации более надежно.

4. Раунд краудсорсинга: раунд краудсорсинга определяет количество пользователей, выбранных моделью для участия в краудсорсинге. Если это первый раунд краудсорсинга, модель игнорирует этот фактор. Если это второй или третий раунд, модель выбирает фиксированное количество пользователей для краудсорсинга и игнорирует все другие факторы. Это связано с тем, что во втором и третьем раундах краудсорсинга модель выбирает только пользователей, удовлетворяющих определенным критериям, для участия в краудсорсинге.

Рисунок 5 иллюстрирует процесс определения количества пользователей для участия в краудсорсинге.

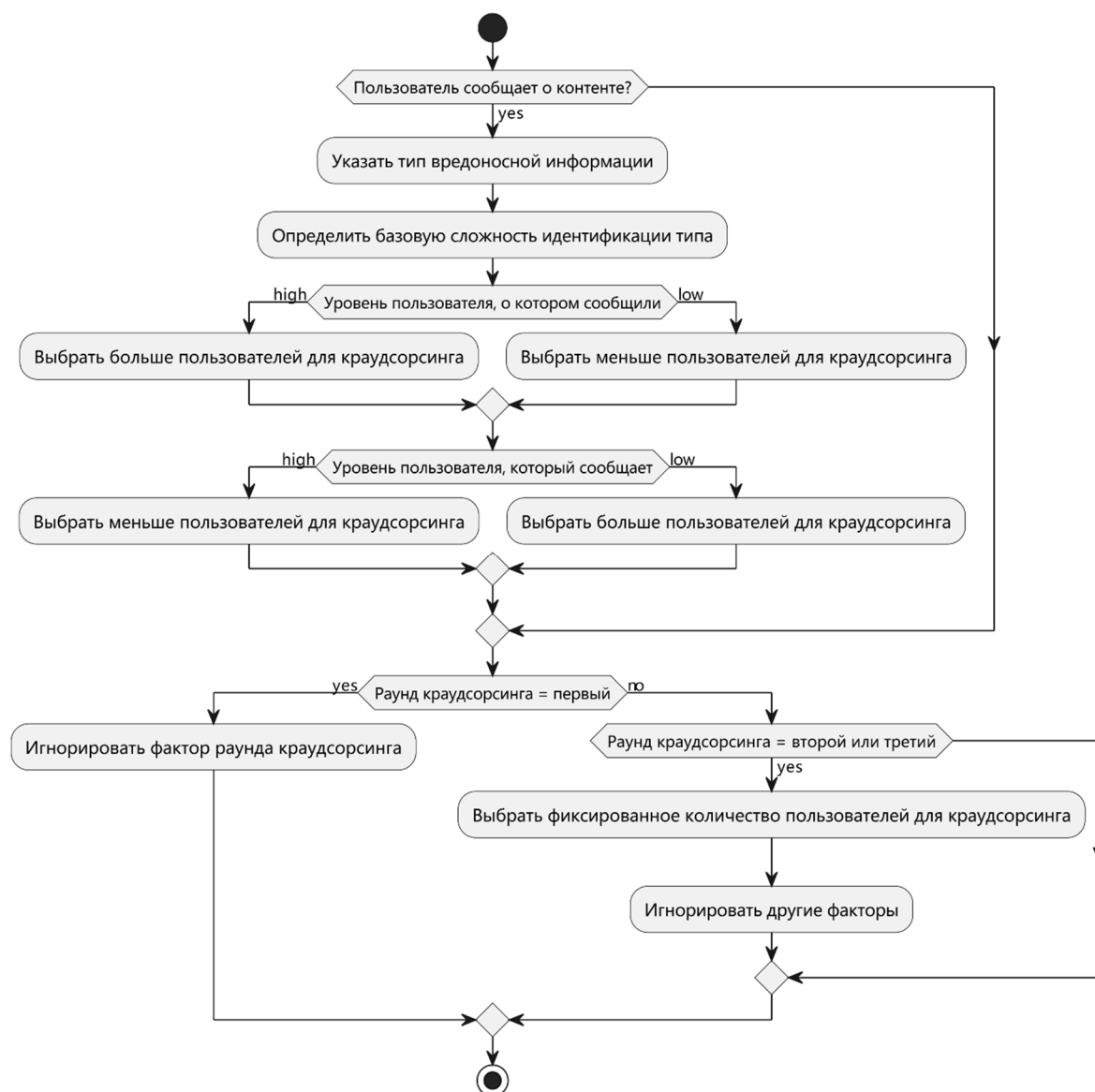


Рисунок 5 – Процесс определения количества пользователей для участия в краудсорсинге

После распознавания необходимого количества пользователей для краудсорсинга модель переходит к выбору тех пользователей, которые должны участвовать в краудсорсинге, учитывая следующие факторы:

1. Раунд краудсорсинга: раунд краудсорсинга напрямую влияет на то, как модель выбирает пользователей для участия в краудсорсинге.

- Если это первый раунд краудсорсинга, пользователи случайным образом выбираются из разных диапазонов уровней пользователей. Модель

разделяет пользователей на три диапазона уровней: пользователи с уровнями между 70 и 80, пользователи с уровнями между 80 и 90 и пользователи с уровнями между 90 и 100. Модель гарантирует, что 70% участников будут из категории уровней между 70 и 80, 20% - из категории уровней между 80 и 90, а 10% - из категории уровней между 90 и 100.

- Если это второй раунд краудсорсинга, для участия в краудсорсинге выбираются только пользователи с уровнем выше 90, и по крайней мере один из пользователей должен иметь роль эксперта или администратора.

- Если это третий раунд краудсорсинга, для участия в краудсорсинге выбираются только пользователи с ролью "эксперт" или "администратор".

На Рисунке 6 представлена схема процесса выбора пользователей для участия в краудсорсинге.

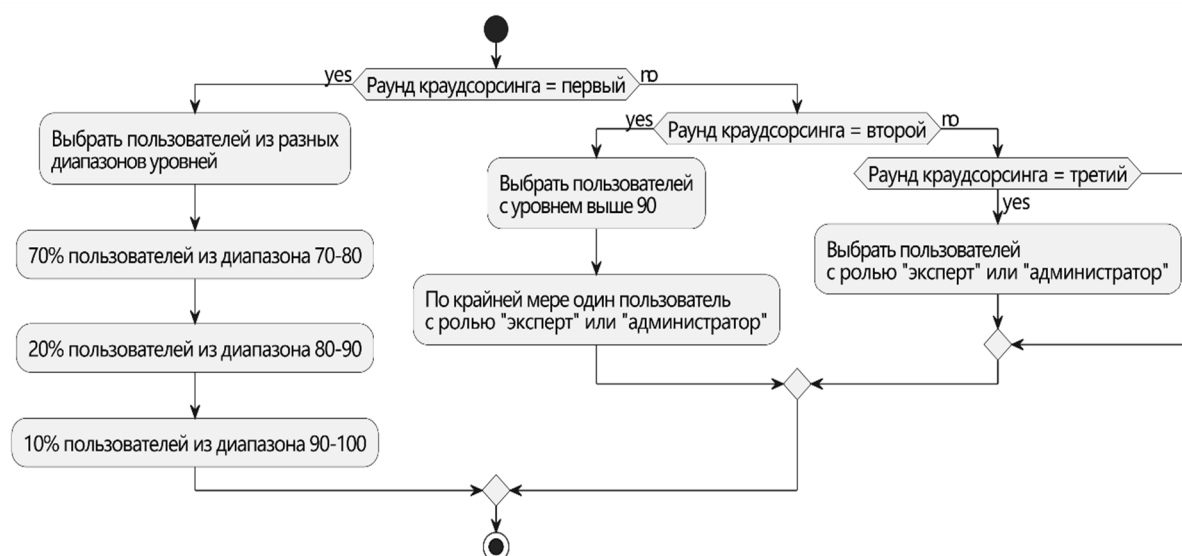


Рисунок 6 – Процесс выбора пользователей для участия в краудсорсинге.

## 2.4 Третья часть модели: Механизм расчета результатов краудсорсинга

Третья часть модели отвечает за расчет результатов краудсорсинга. Ее задача - учитывать голоса пользователей в краудсорсинге и получать предварительный результат краудсорсинга и индекс спорности.

Существует три возможности для результатов голосования в краудсорсинге пользователей:

- -1: Означает возражение против отчета о действительности, что означает, что сообщенный контент, по мнению голосующего, не содержит вредоносной информации, указанной докладчиком.

- 0: Означает воздержание. Когда пользователи не могут сделать суждение, они могут выбрать воздержаться от краудсорсинга. Кроме того, если пользователь не делает никакого суждения после окончания периода краудсорсинга, его голос считается воздержанием.

- 1: Означает согласие с действительностью отчета, что означает, что сообщенный контент действительно содержит вредоносную информацию, указанную докладчиком.

Среди этих трех возможностей Пользователи, которые голосуют за -1 или 1, считаются действительными пользователями краудсорсинга, и их голоса считаются действительными голосами краудсорсинга. Пользователи, которые голосуют за 0, считаются недействительными пользователями краудсорсинга, а их голоса называются недействительными голосами краудсорсинга.

После завершения процесса голосования в формате краудсорсинга, модель проводит взвешенную обработку действительных голосов краудсорсинга и вычисляет предварительный результат краудсорсинга. При вычислении предварительного результата краудсорсинга модель учитывает следующие факторы:

1. Количество недействительных пользователей краудсорсинга: Если количество недействительных пользователей краудсорсинга превышает половину общего числа пользователей краудсорсинга в конкретном краудсорсинге, краудсорсинг считается недействительным.

2. Уровень пользователя действительных пользователей краудсорсинга:  
В краудсорсинге модель назначает разные веса действительным пользователям краудсорсинга на основе их уровней пользователей и использует их для расчета взвешенного результата краудсорсинга.

Формула (1) расчета веса для действительных пользователей голосования:

$$w_u = \frac{l_u^4}{\sum_{i=1}^n l_{u_i}^4} \quad (1)$$

где  $w_u$  – вес пользователя;

$l_u$  – уровень пользователя;

$n$  – количество всех действующих голосующих.

Формула (2) расчета взвешенного результата краудсорсинга:

$$R = \sum_{i=1}^n w_{u_i} \cdot r_{u_i} \quad (2)$$

где  $R$  – взвешенный результат краудсорсинга.

$w_{u_i}$  – веса действующих голосующих пользователей;

$r_{u_i}$  – их соответствующие результаты голосования;

$n$  – количество всех действующих голосующих.

Затем модель рассчитывает значение предварительного результата краудсорсинга на основе взвешенного результата краудсорсинга:

$$P(R) = \begin{cases} 0, & \text{if } -0.25 \leq R \leq 0.25 \\ -1, & \text{if } -1 \leq R < -0.25 \\ 1, & \text{if } 0.25 < R \leq 1 \end{cases} \quad (3)$$



где  $R$  – взвешенный результат краудсорсинга.

Есть три возможности для предварительного результата краудсорсинга:

- -1: Означает возражение против действительности отчета.
- 0: Означает недействительный результат краудсорсинга.
- 1: Означает согласие с действительностью отчета.

Если взвешенный результат краудсорсинга находится в диапазоне от -0.25 до 0.25, предварительный результат краудсорсинга принимается равным 0. Если взвешенный результат краудсорсинга находится между -1 и -0.25, предварительный результат краудсорсинга записывается как -1. Если взвешенный результат краудсорсинга находится между 0.25 и 1, предварительный результат краудсорсинга записывается как 1.

На основе предварительного результата краудсорсинга модель определяет дальнейшие шаги. Если предварительный результат краудсорсинга равен -1 или 1, он считается действительным результатом краудсорсинга. Если предварительный результат краудсорсинга равен 0, он считается недействительным результатом краудсорсинга.

Для действительных результатов краудсорсинга модель переходит к расчету индекса спорности краудсорсинга и определяет надежность предварительного результата краудсорсинга на основе индекса спорности.

Для недействительных результатов краудсорсинга модель рассматривает краудсорсинг как неудачу, и этот раунд краудсорсинга заканчивается, что приводит к началу следующего раунда краудсорсинга для сообщенного контента.

На Рисунке 7 показано, как рассчитываются результаты краудсорсинга.

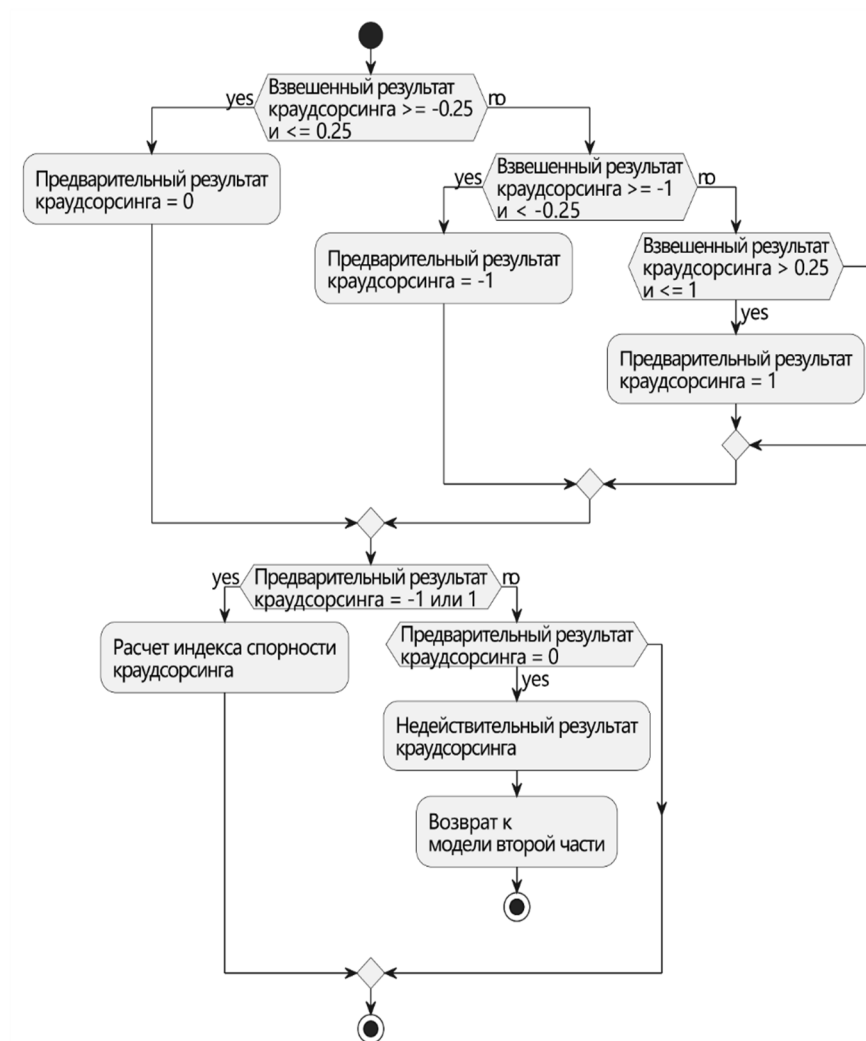


Рисунок 7 – Расчет результатов краудсорсинга.

## 2.5 Четвертая часть модели: Механизм обработки споров в краудсорсинге

Четвертая часть модели отвечает за определение надежности предварительного результата краудсорсинга. Если предварительный результат краудсорсинга является действительным, модель производит расчет индекса спорности.

Индекс спорности представляет степень несогласия между участниками краудсорсинга и является абсолютным значением взвешенного результата краудсорсинга. Меньший индекс спорности указывает на более высокий уровень несогласия между пользователями и нижнюю надежность результата краудсорсинга.

Согласно формуле (4) рассчитывается индекса спорности:

$$D = |R| \quad (4)$$

где  $D$  - индекса спорности;

$R$  – взвешенный результат краудсорсинга.

Исходя из значения индекса спорности, степень несогласия пользователей классифицируется следующим образом:

- Указывает на значительное несогласие среди пользователей, что снижает надежность результата краудсорсинга.
- Индекс спорности между 0.5 и 0.75: указывает на незначительное несогласие среди пользователей, что приводит к повышенной надежности результата краудсорсинга.
- Индекс спорности между 0.75 и 1: указывает на минимальное или почти полное отсутствие несогласия среди пользователей, что приводит к чрезвычайно высокой надежности результата краудсорсинга.

Если заявивший пользователь или сообщивший пользователь не согласны с предварительным результатом краудсорсинга, рассматриваются несколько факторов, чтобы определить, имеет ли пользователь право на апелляцию:

1. Уровень пользователя заявившего или сообщившего пользователя: Если у заявившего пользователя уровень больше 70 и рассчитанный индекс спорности меньше 0.75, и если краудсорсинг находится на первом раунде, то заявившему пользователю предоставляется возможность непосредственно обжаловать. Это связано с тем, что пользователи с более высоким уровнем считаются надежными, и модель более осторожно обрабатывает краудсорсинг, связанный с надежными пользователями. Наоборот, если окончательный

результат краудсорсинга для пользователей высокого уровня согласуется с отчетом, наказание, наложенное на пользователя, будет более суровым.

Если у пользователя нет прямой возможности обжаловать, модель рассчитывает значение, называемое коэффициентом уровня пользователя, с использованием формулы (5):

$$C_u(L) = \begin{cases} 0.8, & \text{if } L < 50 \\ 1, & \text{if } 50 \leq L < 70 \\ 1.2, & \text{if } L \geq 70 \end{cases} \quad (5)$$

где  $L$  - уровень пользователя.

2. Тип заявленного контента: модель рассчитывает значение, называемое коэффициентом типа, исходя из типа заявленного контента. Коэффициент типа обозначается как  $C_t$ .

3. Индекс спорности: Индекс спорности используется при расчете возможности апелляции.

$$C_d(D) = 1 - D \quad (6)$$

где  $D$  - индекса спорности.

4. Раунды краудсорсинга: модель рассчитывает значение, называемое коэффициентом раунда, с использованием следующей формулы:

$$C_r(RO) = \begin{cases} 1, & \text{if } RO = 1 \\ 0.8, & \text{if } RO = 2 \\ 0, & \text{if } RO > 2 \end{cases} \quad (7)$$

где  $RO$  - Раунды краудсорсинга.

Исходя из рассчитанных коэффициента уровня пользователя, коэффициента типа, индекса спорности и коэффициента раунда, модель определяет, имеет ли пользователь право на апелляцию, используя следующую формулу:

$$P = C_u \cdot C_t \cdot C_d \cdot C_r \quad (8)$$

где  $C_u$  - коэффициента уровня пользователя;

$C_t$  - коэффициента типа;

$C_d$  – коэффициента индекса спорности;

$C_r$  - коэффициента раунда.

Если значение, рассчитанное с использованием этой формулы, ниже определенного порога, пользователю не разрешается подавать апелляцию. Если рассчитанное значение превышает определенный порог, пользователю разрешается подавать апелляцию.

Если модель позволяет пользователям обжаловать предварительный результат краудсорсинга и пользователь подал апелляцию, модель начинает новый раунд краудсорсинга для заявленного контента.

Если пользователь не обжаловал или не успел обжаловать в указанный срок, модель записывает предварительный результат краудсорсинга как окончательный результат краудсорсинга. На этом этапе модель присваивает баллы краудсорсинга участвующим пользователям на основе окончательного результата краудсорсинга и принимает соответствующие меры по отношению к заявленному контенту и заявившему пользователю.

Если модель определяет, что подача апелляции не разрешена на основе рассчитанного результата, модель записывает предварительный результат краудсорсинга этого раунда как окончательный результат краудсорсинга. На

этом этапе модель присваивает баллы краудсорсинга участвующим пользователям на основе окончательного результата краудсорсинга и принимает соответствующие меры по отношению к заявленному контенту и заявившему пользователю.

Окончательный результат краудсорсинга, полученный моделью, может быть одним из следующих вариантов:

- -1: Противостоит действительности отчета, указывая на то, что заявленный контент не содержит предполагаемой вредоносной информации.
- 1: Поддерживает действительность отчета, указывая на то, что заявленный контент действительно содержит предполагаемую вредоносную информацию.

Схема процесса этой части модели показана на рисунке 8.

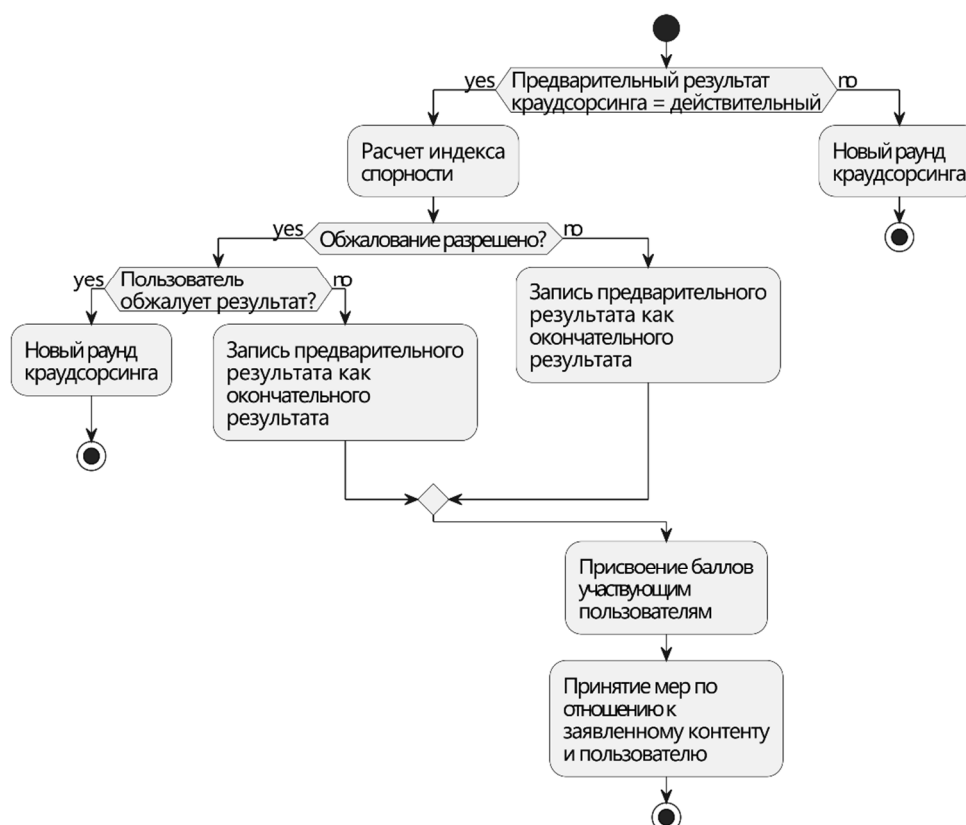


Рисунок 8 – Механизм обработки споров в краудсорсинге

## **2.6 Пятая часть модели: Механизм обработки сообщенного контента**

Эта часть модели отвечает за обработку сообщенного контента на основе окончательного результата краудсорсинга.

Если окончательный результат краудсорсинга противоречит действительности отчета, предпринимаются следующие действия:

- Если новые пользователи сообщают о том же контенте и их уровень пользователя превышает как уровень первоначально сообщившего пользователя, так и уровень пользователя сообщенного контента, для этого контента инициируется новый раунд краудсорсинга.

- Если эксперт или администратор сообщает о том же контенте, инициируется новый раунд краудсорсинга для этого контента.

- Контент может пройти не более трех раундов краудсорсинга. После третьего раунда краудсорсинга, если окончательный результат краудсорсинга по-прежнему противоречит действительности отчета, больше не принимаются отчеты о этом контенте.

Для сообщенного контента:

Если окончательный результат краудсорсинга подтверждает действительность отчета, модель рассчитывает окончательный балл наказания пользователя на основе следующих факторов:

1. Уровень пользователя, о котором сообщается: модель рассчитывает базовое значение наказания на основе уровня пользователя. Для пользователей с уровнем пользователя между 50 и 70, их базовое значение наказания составляет -10 (текущий уровень пользователя минус 10). Для пользователей с уровнем пользователя ниже 50, их базовое значение наказания составляет -20. Для пользователей с уровнем пользователя выше 70, их базовое значение наказания составляет -20.

2. Тип сообщенного контента: модель рассчитывает первый коэффициент наказания на основе типа сообщенного контента. Для вредоносного контента с оскорбительным языком коэффициент наказания составляет 1. Для вредоносного контента с ложной информацией коэффициент наказания составляет 2. Для вредоносного контента, содержащего речь ненависти, коэффициент наказания составляет 2,5. Для вредоносного контента, связанного с мошенничеством, коэффициент наказания составляет 5.

3. Подал ли пользователь, о котором сообщается, апелляцию: модель рассчитывает второй коэффициент наказания на основе того, подал ли пользователь апелляцию. Если пользователь не подал апелляцию, коэффициент наказания составляет 1. Если пользователь подал апелляцию, коэффициент наказания составляет 1,1.

Модель рассчитывает окончательный балл наказания пользователя с использованием следующей формулы:

$$S = F_1 \cdot F_2 \cdot B \quad (9)$$

где  $S$  - окончательный балл наказания;

$F_1$  - коэффициент наказания 1;

$F_2$  - коэффициент наказания 2;

$B$  - базовое значение наказания.

Схема процесса этой части модели показана на рисунке 9.



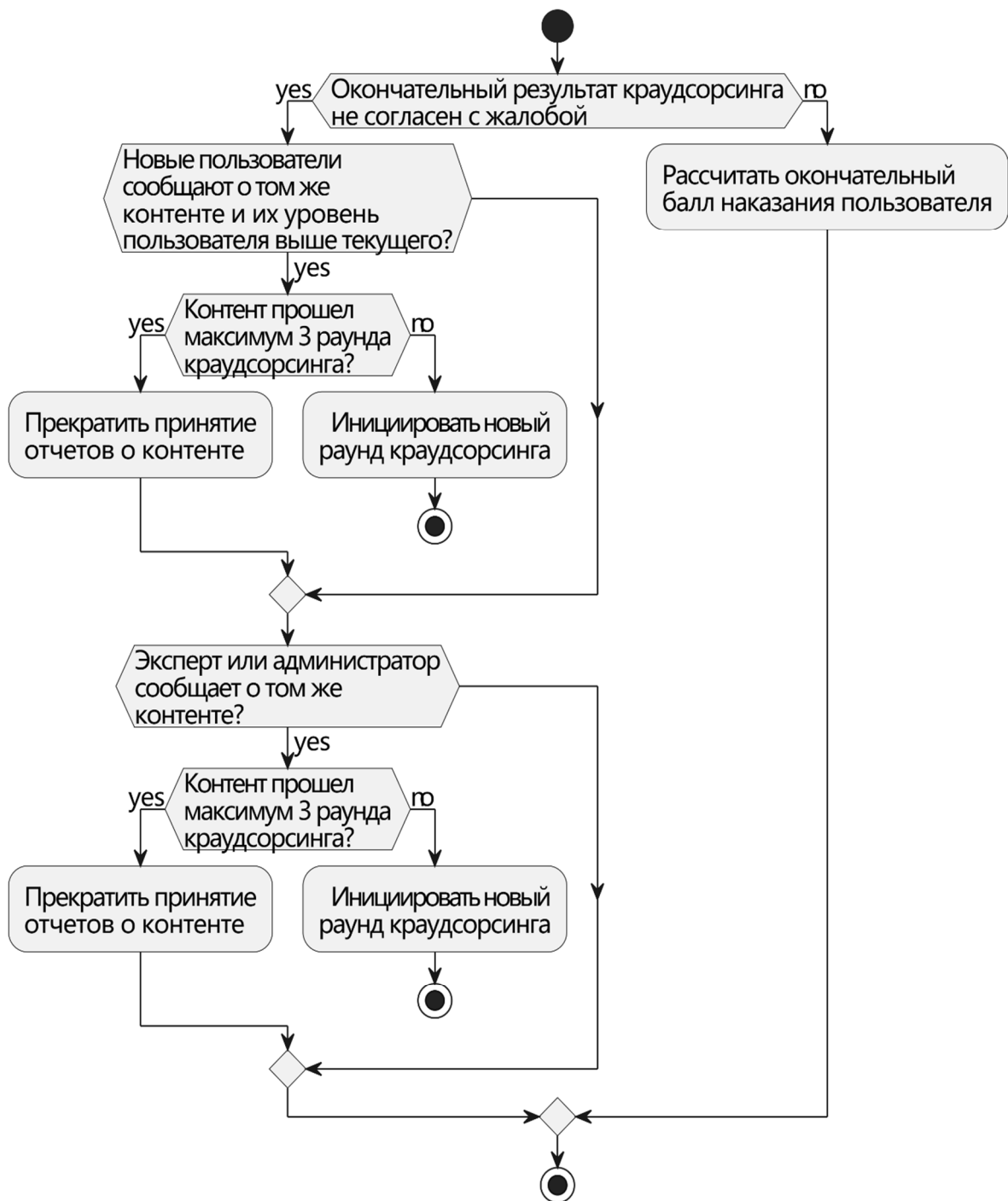


Рисунок 9 – Механизм обработки сообщенного контента

## 2.7 Шестая часть модели: Механизм расчета баллов краудсорсинга

Эта часть модели в основном отвечает за расчет баллов краудсорсинга пользователей, которые участвовали в процессе краудсорсинга.

Для пользователей, которые считаются недействительными пользователями краудсорсинга, то есть они не участвовали или воздержались от голосования, их балл краудсорсинга устанавливается равным 0.

Для пользователей, которые являются действительными пользователями краудсорсинга и чье голосование по краудсорсингу совпадает с окончательным результатом краудсорсинга, они получают положительный балл краудсорсинга.

Для пользователей, которые являются действительными пользователями краудсорсинга, но чье голосование по краудсорсингу не совпадает с окончательным результатом краудсорсинга, они получают отрицательный балл краудсорсинга.

При расчете баллов краудсорсинга модель учитывает следующие факторы:

1. Коэффициент несогласия: Если коэффициент несогласия превышает 0,75, первый коэффициент устанавливается равным 1. Если коэффициент несогласия составляет от 0,5 до 0,75, первый коэффициент устанавливается равным 2. Если коэффициент несогласия составляет от 0,25 до 0,5, первый коэффициент устанавливается равным 3.

2. Уровень пользователя: Если уровень пользователя составляет от 90 до 100, второй коэффициент устанавливается равным 1. Если уровень пользователя составляет от 80 до 90, второй коэффициент устанавливается равным 1.5. Если уровень пользователя составляет от 70 до 80, второй коэффициент устанавливается равным 2.

Для положительных баллов краудсорсинга используется следующая формула для расчета:

$$C_{s_{final}} = K_1 \cdot K_2 \cdot C_{s_{original}} \quad (10)$$

где  $C_{sfinal}$  - балл краудсорсинга;

$C_{soriginal}$  - балл краудсорсинга;

$K_1$  - первый коэффициент;

$K_2$  - второй коэффициент.

Для отрицательных баллов краудсорсинга используется следующая формула для расчета:

$$C_{sfinal} = (K_{1max} - K_1) \cdot (K_{2max} - K_2) \cdot C_{soriginal} \quad (11)$$

где  $C_{sfinal}$  - балл краудсорсинга;

$C_{soriginal}$  - балл краудсорсинга;

$K_{1max}$  - максимальное значение, которое может принять коэффициент  $K_1$ ;

$K_1$  - первый коэффициент;

$K_{2max}$  - максимальное значение, которое может принять коэффициент  $K_2$ ;

$K_2$  - второй коэффициент.

Схема процесса этой части модели показана на рисунке 10.

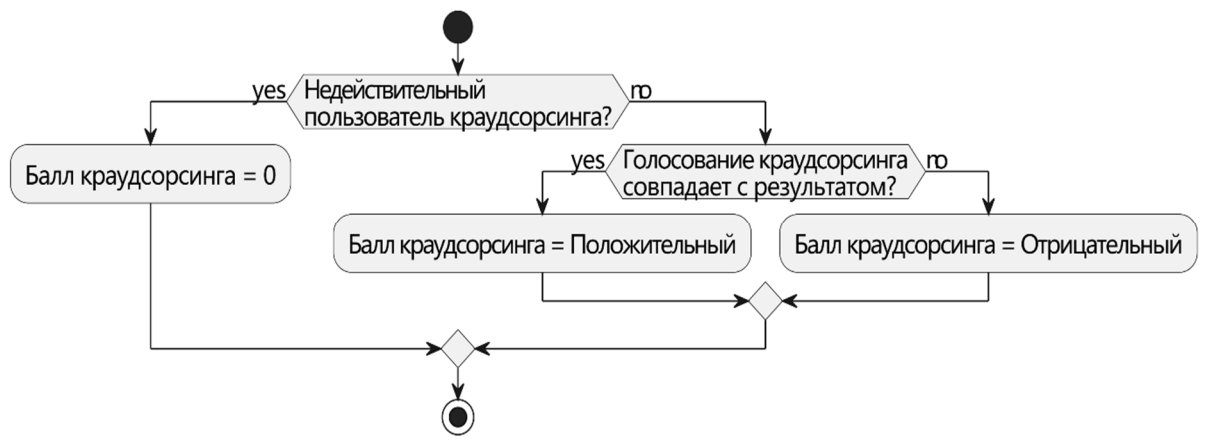


Рисунок 10 – Механизм расчета баллов краудсорсинга

### **3 Валидация с помощью симуляции**

#### **3.1 Дизайн набора данных для симуляции**

Поскольку модель в этом исследовании требует использования большого количества пользовательских данных и информации о жалобах, которые содержат множество чувствительных данных, социальные сетевые платформы не предоставляют реальный набор данных, необходимый для проверки модели, чтобы обеспечить защиту конфиденциальности пользователей и соблюдение законов и правил. Следовательно, это исследование использует предустановленные параметры для создания набора данных для симуляции и использует этот набор данных и симуляционную модель для проверки эффективности модели.

При разработке моделей симуляции и создании наборов данных для симуляции данное исследование выбрало Python в качестве основного языка программирования. Python является одним из наиболее популярных языков программирования в настоящее время. В этом исследовании также было выбрано использование Anaconda и Jupyter для создания рабочей среды.

Наборы данных для симуляции включают следующее:

1. Набор данных о пользователях: этот набор данных содержит специфическую информацию о пользователях.
2. Набор данных о жалобах: этот набор данных содержит специфическую информацию о жалобах.

Для набора данных о пользователях включены следующие данные:

1. `user_id`: идентификатор пользователя. Каждый пользователь в наборе данных имеет уникальный идентификатор, где каждый идентификатор представляет пользователя. Идентификатор пользователя – это целое число, начинающееся с 0.
2. `user_level`: уровень пользователя. У каждого пользователя в наборе данных есть свой уровень пользователя, который представляет собой

надежность пользователя в модели. Пользователи с более высоким уровнем пользователя считаются более надежными в модели. Уровень пользователя - это число с плавающей запятой, варьирующееся от 20 до 100.

3. `participation_probability`: вероятность участия пользователя. Вероятность участия - это число с плавающей запятой, варьирующееся от 0 до 1, указывающее вероятность участия пользователя в процессе краудсорсинга.

4. `correct_probability`: вероятность правильности пользователя. Вероятность правильности - это число с плавающей запятой, варьирующееся от 0 до 1, указывающее вероятность правильного суждения пользователя.

5. `selected_times`: количество раз, когда пользователь был выбран для участия в краудсорсинге. Количество выбранных раз – это целое число, начинающееся с 0.

6. `user_role`: флаг роли пользователя. Флаг роли пользователя имеет только два значения: 0 или 1. Если значение равно 0, это представляет обычного пользователя. Если значение равно 1, это представляет администратора или экспертного пользователя.

7. `user_status`: флаг состояния пользователя. Флаг состояния пользователя имеет только два значения: 0 или 1. Если значение равно 0, это представляет пользователя в нормальном состоянии. Если значение равно 1, это представляет пользователя в ограниченном состоянии.

Для набора данных о жалобах включены следующие данные:

1. `report_id`: идентификатор сообщенного контента. Каждому сообщенному контенту присваивается уникальный идентификатор для распознавания сообщенного контента. Идентификатор сообщения – это целое число, начинающееся с 0.

2. `report_type`: тип сообщенного контента. Тип сообщения имеет три значения: 0, 1, 2. Это значение представляет уровень сложности распознавания сообщенного контента. 0 представляет относительно легкое

суждение, 1 представляет умеренно сложное суждение, а 2 представляет чрезвычайно сложное суждение.

3. `difficulty_factor`: коэффициент сложности правильного суждения для пользователя. Это значение - это число с плавающей запятой, варьирующееся от 0.5 до 1. Чем ниже коэффициент сложности, тем больше сложность для пользователя сделать правильное суждение, указывающее на меньшую вероятность сделать правильное суждение. Коэффициент сложности положительно коррелирует с типом сообщенного контента.

4. `reported_user_id`: идентификатор пользователя-создателя сообщенного контента.

5. `report_user_id`: идентификатор пользователя, который сообщил контент.

6. `judgment_rounds`: текущий раунд краудсорсинга. Это значение имеет только три значения: 1, 2, 3, представляющих соответствующий раунд краудсорсинга.

7. `participant_count`: количество пользователей, участвующих в текущем краудсорсинге. Для сообщенного контента модель определяет количество пользователей, участвующих в этом краудсорсинге. Это значение - положительное целое число.

8. `judgment_score_weighted`: взвешенный результат краудсорсинга текущего раунда. Это значение рассчитывается моделью на основе результатов голосования участвующих пользователей. Это число с плавающей запятой, варьирующееся от -1 до 1.

9. `dispute_coefficient`: коэффициент спорности текущего краудсорсинга. Коэффициент спорности получается моделью на основе значения взвешенного результата краудсорсинга. Это число с плавающей запятой, варьирующееся от 0 до 1. Меньшее значение указывает на более высокий уровень спорности и ниже надежность результата краудсорсинга.

10. `judgment_result`: результат текущего краудсорсинга. Это значение рассчитывается моделью на основе взвешенного результата краудсорсинга. Оно имеет только три значения: -1, 0, 1.

11. `final_judgment_sign`: финальный флаг краудсорсинга. Это значение имеет только два значения: 0 или 1. Если значение равно 0, это указывает на предварительный результат краудсорсинга. Если значение равно 1, это указывает на финальный результат краудсорсинга.

12. `process_result_status`: статус обработки. Это значение представляет действия, предпринятые моделью для сообщенного контента. Оно имеет только три значения: -1, 0, 1. Если значение равно -1, это означает, что модель не предпринимает никаких действий по отношению к сообщенному контенту. Если значение равно 0, это означает, что модель временно не обрабатывает сообщенный контент. Если значение равно 1, это означает, что модель приняла соответствующие действия по отношению к сообщенному контенту.

### **3.2 Проектирование симуляционной модели**

Целью проектирования симуляционной модели является проверка эффективности предложенной модели распознавания вредоносной информации на основе краудсорсинга пользователей в данном исследовании. Поэтому симуляционная модель реализует только основные функции теоретической модели.

Симуляционная модель реализует следующие функции:

1. Функция `generate_user_data`: вызов этой функции позволяет симуляционной модели генерировать заданное количество пользовательских данных. Пользовательские данные генерируются в соответствии с предварительно заданными параметрами. Эти пользовательские данные будут использоваться в качестве симуляционного набора данных.

Сначала функция генерирует заданное количество уникальных идентификаторов пользователей (`user_id`). Затем на основе предварительно



заданных параметров генерируются уровни пользователей (`user_levels`), которые соответствуют нормальному распределению и находятся в диапазоне от 20 до 100. Затем генерируются вероятности участия (`participation_probabilities`), вероятности правильного ответа (`correct_probabilities`) и выбранные времена (`selected_times`), которые также следуют нормальному распределению. Также генерируются выбранные времена (`selected_times`), которые соответствуют предварительно заданным параметрам и следуют нормальному распределению.

В то же время функция гарантирует положительную корреляцию между уровнями пользователей (`user_levels`) и вероятностями участия (`participation_probabilities`), вероятностями правильного ответа (`correct_probabilities`) и выбранными временами (`selected_times`). Другими словами, пользователи с более высокими уровнями пользователей имеют более высокие вероятности участия, вероятности правильного ответа и выбранные времена.

Наконец, значения роли пользователя (`user_role`) равны 1 для данных с `user_levels`, равным 100, и `participation_probabilities`, превышающими 0.95, а для остальных данных значения роли пользователя (`user_role`) равны 0. Кроме того, все пользователи получают значения статуса пользователя (`user_status`) равные 0.

2. функция `generate_report_data`: При вызове этой функции симуляционная модель генерирует заданное количество данных о жалобах. Данные о жалобах генерируются в соответствии с предопределенными параметрами. Эти данные о жалобах будут использоваться в качестве симуляционного набора данных.

Сначала функция генерирует заданное количество уникальных идентификаторов жалоб (`report_id`). Затем для каждого `report_id` на основе предварительно заданных параметров генерируются тип жалобы (`report_type`) и коэффициент сложности (`difficulty_factor`).

Затем для `reported_user_id` и `report_user_id` симуляционная модель случайным образом выбирает определенное количество `user_id` из набора пользовательских данных на основе разных `user_levels`.

Реализация происходит следующим образом:

Сначала пользовательский набор данных разделяется на три поднабора в соответствии с уровнем пользователя:

- `user_ids_reported`: содержит всех пользователей с `user_level` менее 70.

В этом поднаборе не все пользователи имеют право на подачу жалобы, так как модель ограничивает пользователей с `user_level` менее 60 от подачи жалоб.

- `user_ids_report`: содержит всех пользователей с `user_level` от 60 до 70.

В этом поднаборе все пользователи имеют право на подачу жалобы.

- `user_ids_high`: содержит всех пользователей с `user_level` 70 и выше. В этом поднаборе все пользователи имеют право на подачу жалобы.

Во всем наборе данных о жалобах 80% `reported_user_id` случайным образом выбираются из `user_ids_reported user_id`, а оставшиеся 20% `reported_user_id` случайным образом выбираются из `user_ids_high user_id`.

Аналогично, 80% `report_user_id` случайным образом выбираются из `user_ids_high user_id`, а оставшиеся 20% `report_user_id` случайным образом выбираются из `user_ids_report user_id`.

После завершения этого процесса во всем наборе данных о жалобах 80% пользователей, на которых поданы жалобы, имеют `user_level` менее 70, а 20% пользователей, на которых поданы жалобы, имеют `user_level` 70 и выше. Аналогично, 80% пользователей, подающих жалобы, имеют `user_level` 70 и выше, а 20% пользователей, подающих жалобы, имеют `user_level` от 60 до 70.

Наконец, функция устанавливает значение `judgment_rounds` равным 1 для всех данных в наборе данных о жалобах и устанавливает значения `participant_count`, `judgment_score_weighted`, `dispute_coefficient`, `judgment_result`, `final_judgment_sign` и `process_result_status` равными 0.

3. Функция `get_participant_count_parallel`: При вызове этой функции вычисляется количество участников, принимающих участие в краудсорсинге пользователей. Она вычисляет значение `participant_count` в наборе данных о жалобах. Эта функция реализует функциональность вычисления количества участников, принимающих участие в краудсорсинге пользователей, что является частью теоретической модели : Механизма выбора краудсорсинга.

4. Функция `select_users`: При вызове этой функции выбираются пользователи, участвующие в краудсорсинге пользователей. Эта функция реализует функциональность выбора определенных пользователей для участия в процессе краудсорсинга пользователей. Это также является частью теоретической модели : Механизма выбора краудсорсинга.

5. функция `generate_judgement_results_parallel`: При вызове этой функции генерируются результаты голосования пользователей, участвующих в краудсорсинге оценки. Результаты оценки пользователя зависят от параметров участия (`participation_probabilities`), параметров правильных ответов (`correct_probabilities`), роли пользователя (`user_role`) и коэффициента сложности контента, на который подана жалоба (`difficulty_factor`).

Сначала функция случайным образом генерирует данные, представляющие участие пользователя в краудсорсинге, на основе параметров участия (`participation_probabilities`). Результат голосования пользователя генерируется случайным образом между 0 и 1, с вероятностью генерации 1, равной параметрам участия (`participation_probabilities`) пользователя. Если результат голосования пользователя равен 0, это означает, что пользователь не участвовал в этом краудсорсинге и считается недействительным участником.

Затем, если результат голосования пользователя равен 1, определяется, будет ли пользователь подвержен влиянию коэффициента сложности (`difficulty_factor`), в зависимости от его роли пользователя (`user_role`). Если роль пользователя равна 0, его правильность оценки определяется как

произведение его собственных параметров правильности (`correct_probabilities`) на коэффициент сложности (`difficulty_factor`). Если роль пользователя равна 1, его правильность оценки определяется его собственными параметрами правильности (`correct_probabilities`).

Наконец, на основе правильности пользователя, результат голосования пользователя генерируется случайным образом между -1 и 1, с вероятностью генерации 1, равной правильности пользователя.

Если результат голосования пользователя равен 1, это означает, что пользователь участвовал в краудсорсинге и дал правильную выбор.

Если результат голосования пользователя равен 0, это означает, что пользователь не участвовал в краудсорсинге.

Если результат голосования пользователя равен -1, это означает, что пользователь участвовал в краудсорсинге, но дал неправильную выбор.

С помощью такого подхода можно имитировать реальные результаты голосования пользователей. Участие в краудсорсинге зависит от параметров участия пользователя, правильность определяется его параметрами правильности, ролью пользователя и сложностью контента, на который подана жалоба.

6. функция `calculate_weighted_scores_parallel`: При вызове этой функции производится расчет взвешенных оценок и коэффициента спора для результатов краудсорсинга. Это включает расчет взвешенной оценки (`judgment_score_weighted`) и коэффициента спора (`dispute_coefficient`). Данная функция реализует механизм расчета результатов краудсорсинга, предусмотренный теоретической моделью.

7. функция `calculation_results`: При вызове этой функции производится расчет предварительных результатов краудсорсинга. Данная функция реализует механизм расчета результатов краудсорсинга, предусмотренный теоретической моделью.

8. функция `handle_disputes`: При вызове этой функции осуществляется обработка споров и определение возможности пользователей обратиться. Данная функция реализует механизм обработки споров в краудсорсинге, предусмотренный теоретической моделью.

9. функция `record_result`: При вызове этой функции происходит разделение обработанных результатов и результатов, требующих дальнейшей обработки.

Для результатов, требующих дальнейшей обработки, производится переход к следующему раунду краудсорсинга.

После завершения трех раундов краудсорсинга, модель симуляции завершается.

### **3.3 Анализ результатов симуляции**

Для набора данных симуляции были использованы следующие предустановленные параметры:

1. Для набора данных пользователя, симуляция установила количество пользователей в 100,000.

- Для `user_levels` предполагалось, что они следуют нормальному распределению со средним значением 70 и стандартным отклонением 10.

- Для `participation_probabilities` предполагалось, что они следуют нормальному распределению со средним значением 0.7 и стандартным отклонением 0.2.

- Для `correct_probabilities` предполагалось, что они следуют нормальному распределению со средним значением 0.8 и стандартным отклонением 0.1.

- Для `selected_times` предполагалось, что они следуют нормальному распределению со средним значением 1000 и стандартным отклонением 100.

- Были сделаны некоторые корректировки экстремальных значений данных.

Предустановленные параметры для набора данных пользователя приведены в следующей таблице 2.

Таблица 2 – Предустановленные параметры для набора данных пользователя

Параметр	Среднее значение	Стандартное отклонение
user_levels	70	10
participation_probabilities	0.7	0.2
correct_probabilities	0.8	0.1
selected_times	1000	100

2. Для набора данных отчетов, симуляция установила количество отчетов в 10,000.

- Для report\_types предполагалось, что распределение значений 0, 1 и 2 будет соответственно 0.7, 0.25 и 0.05.

- Для данных с report\_types, равными 0, предполагалось, что difficulty\_factor следует нормальному распределению со средним значением 0.95 и стандартным отклонением 0.025, но гарантировалось, что он будет между 0.9 и 1.

- Для данных с report\_types, равными 1, предполагалось, что difficulty\_factor следует нормальному распределению со средним значением 0.8 и стандартным отклонением 0.25, но гарантировалось, что он будет между 0.7 и 0.9.

- Для данных с report\_types, равными 2, предполагалось, что difficulty\_factor следует нормальному распределению со средним значением 0.6 и стандартным отклонением 0.25, но гарантировалось, что он будет между 0.5 и 0.7.

Предустановленные параметры для набора данных отчетов приведены в следующей таблице 3.

Таблица 3 – Предустановленные параметры для набора данных отчетов

Параметр	Доля	Среднее значение	Стандартное отклонение	Ограничения
difficulty_factor (report_types = 0)	0.7	0.95	0.025	0.9 - 1
difficulty_factor (report_types = 1)	0.25	0.8	0.25	0.7 - 0.9
difficulty_factor (report_types = 2)	0.05	0.6	0.25	0.5 - 0.7

При этих предустановленных параметрах результаты обработки этих 10,000 сообщенных контентов следующие:

Таблица 4 – Результаты распознавания модели симуляции - набор данных для симуляции

Тип	Общее количество	Количество правильных распознаваний	Количество неправильных распознаваний
report_types = 0	6962	6958	4
report_types = 1	2518	2506	12
report_types = 2	520	500	20
report_types = 0,1,2	10000	9964	36

Рассчитывается:

- Общая точность модели симуляции составляет 99.64%.

- Для сообщений, которые относительно легко определить, точность модели симуляции составляет 99.94%.

- Для сообщений, которые относительно сложно определить, точность модели симуляции составляет 99.52%.

- Для сообщений, которые очень сложно определить, точность модели симуляции составляет 96.15%.

### **3.4 Сравнение моделей**

#### **1. Сравнение различных методов распознавания**

Различные методы распознавания вредоносной информации имеют свои преимущества и недостатки. В этом разделе мы оценим и сравним разные методы распознавания на основе следующих параметров:

- Точность: Точность относится к отношению правильного распознавания вредоносной информации к общему количеству информации. Высокая точность означает, что метод может эффективно отличать вредоносную информацию от нормальной, снижая вероятность ложных срабатываний.

- Скорость распознавания: Скорость распознавания относится к времени, необходимому для вынесения суждения о куске информации. Быстрая скорость распознавания имеет решающее значение для реального времени и масштабных приложений, позволяя быстро обнаружить и предотвратить распространение вредоносной информации.

- Стоимость распознавания: Стоимость распознавания включает в себя ресурсы и расходы, необходимые для метода распознавания. В это входят аппаратные ресурсы, затраты на разработку и поддержку программного обеспечения, время и ресурсы на обучение модели, а также возможные



затраты на рабочую силу. Низкая стоимость распознавания означает, что метод относительно экономичен.

- **Масштабируемость:** Масштабируемость относится к способности метода распознавания адаптироваться к новой вредоносной информации. Это мера того, насколько легко метод может расширяться для обработки новых типов и паттернов вредоносной информации. Метод с хорошей масштабируемостью может быстро реагировать на новую вредоносную информацию.

- **Устойчивость к обходу:** Устойчивость к обходу относится к способности правильно идентифицировать вредоносную информацию, даже когда используются техники обхода. Это предотвращает обход детекции вредоносной информации с помощью техник обхода. Высокая устойчивость к обходу означает, что метод может эффективно обнаруживать вредоносную информацию, используя различные техники обхода, что затрудняет обход.

Для модели распознавания вредоносной информации, созданной в этом исследовании, после тестирования с имитационными данными и моделями, ее уровень точности относительно высок. Поскольку модель основана на участии реальных пользователей, ее масштабируемость очень высока, поскольку она может легко идентифицировать новые типы и паттерны вредоносной информации. В то же время модель также обладает высоким уровнем устойчивости к обходу. Ресурсы аппаратного обеспечения и ресурсы разработки программного обеспечения, необходимые для модели, очень низки, но из-за участия пользователей и администраторов общая стоимость распознавания находится на среднем уровне. Однако, поскольку пользователи, участвующие в совместном суждении, могут потратить некоторое время на обработку суждений, скорость распознавания низкая. В общем, модель распознавания вредоносной информации, созданная в этом исследовании, имеет очень высокую точность, высокую масштабируемость, низкую

устойчивость к обходу, относительно высокую стоимость распознавания и низкую скорость распознавания. Следовательно, эта модель больше подходит для вторичного распознавания вредоносной информации, т.е. проведения второго осмотра контента, который с высокой вероятностью может быть вредоносным, с целью повышения точности.

Для метода распознавания на основе фильтрации по ключевым словам его точность низкая и подвержена ложным срабатываниям. Ключевые слова могут быть легко обойдены, и они не влияют на невредоносный контент, который не содержит ключевые слова. Поэтому масштабируемость метода на основе фильтрации по ключевым словам очень низкая, так же как и его устойчивость к обходу. Однако фильтрация по ключевым словам является самым простым методом распознавания для разработки и имеет очень низкие затраты на разработку по сравнению с другими методами. Таким образом, стоимость распознавания для метода на основе фильтрации по ключевым словам очень низкая. Кроме того, фильтрация по ключевым словам может быстро идентифицировать вредоносную информацию, поэтому скорость распознавания очень высокая.

Для системы ручной модерации, ее точность очень высока, потому что подозрительный контент, сообщенный пользователями, проверяется обученными профессиональными рецензентами. Этот метод имеет высокий уровень масштабируемости и устойчивости к обходу. Однако стоимость труда человека обычно высока, и по мере увеличения числа пользователей и вредоносной информации, количество профессиональных рецензентов также увеличивается. Накопленные отчеты могут не быть обработаны вовремя, что приводит к высокой стоимости распознавания и медленной скорости распознавания для этого метода распознавания.

При использовании методов машинного обучения и глубокого обучения для распознавания вредоносной информации, точность этого метода зависит от большого количества тренировочных данных и самой модели. Только

сложные нейронные сети, обученные на качественных обучающих данных, могут достичь высокой точности. Кроме того, точность распознавания сильно варьируется для различных типов вредоносной информации. Этот метод имеет низкую масштабируемость и устойчивость к обходу. Для распознавания новых форм вредоносной информации требуется собрать большое количество данных для аннотации, а затем обучить модель с этими данными для распознавания новых форм вредоносной информации. Создатели вредоносного контента могут использовать различные техники обхода для обхода этого метода детекции. Например, пользователи, пытающиеся поделиться графическими изображениями, могут изменить цвет крови на изображении, чтобы обойти обнаружение. Кроме того, стоимость распознавания этого метода относительно высока, так как ресурсы аппаратного обеспечения во время обучения модели, ресурсы программного обеспечения для разработки модели, и затраты на рабочую силу для разметки обучающих данных являются дорогостоящими. В отношении скорости распознавания, этот метод медленнее, чем фильтрация по ключевым словам, но гораздо быстрее, чем методы распознавания, основанные преимущественно на ручном.

Оценка четырех методов распознавания по различным параметрам представлена в таблице 5:

Таблица 5 – Оценка четырех методов распознавания по различным параметрам представлена

Метод распознава ния	Точность	Скорость распознава ния	Стоимость распознава ния	Масштабир уемость	Устойчивос ть к обходу
Фильтраци я по ключевым словам	Низкая	Очень высокая	Очень низкая	Очень низкая	Очень низкая

Окончание таблицы 5

Метод распознава ния	Точность	Скорость распознава ния	Стоимость распознава ния	Масштабир уемость	Устойчивос ть к обходу
Машинное обучение и глубокое обучение	Высокая	Высокая	Высокая	Низкая	Низкая
Модель краудсорси нга пользовател ей	Очень высокая	Низкая	Высокая	Очень высокая	Очень высокая
Система ручной модерации	Очень высокая	Очень низкая	Очень высокая	Очень высокая	Очень высокая

На основе приведенной выше таблицы можно заметить, что как методы машинного и глубокого обучения, так и метод фильтрации по ключевым словам имеют достаточно быструю скорость распознавания, что обусловлено способностью машин быстро обрабатывать задачи. Быстрая скорость распознавания делает их подходящими для быстрого обнаружения вредоносной информации. Хотя методы машинного и глубокого обучения имеют более высокие затраты на распознавание по сравнению с методом фильтрации по ключевым словам, они значительно повышают точность и способны справляться с более сложными задачами.

С другой стороны, модель краудсорсинга пользователей и Система ручной модерации имеют более низкую скорость распознавания. Однако они

проявляют очень высокую точность, масштабируемость и устойчивость к обходу, используя преимущества человеческого распознавания. По сравнению с системой ручной модерации, модель краудсорсинга пользователей жертвует некоторой точностью для снижения затрат на распознавание и увеличения скорости распознавания. Это обусловлено тем, что модераторы обычно проходят специальную подготовку, что позволяет им более точно распознавать вредоносную информацию. Однако стоимость найма модераторов является значительной, особенно для крупных социальных сетей, которым требуется большое количество профессиональных модераторов. Важно отметить, что подход краудсорсинга гарантирует, что определение и интерпретация вредоносной информации не контролируются платформой или небольшой группой лиц. Таким образом, в сценариях, требующих точного распознавания вредоносной информации, модель краудсорсинга пользователей, предложенная в этом исследовании, является более подходящей.

Для большинства социальных сетей эти методы распознавания совместимы друг с другом. Например, крупная социальная сеть может использовать фильтрацию по ключевым словам для быстрого распознавания легко узнаваемой вредоносной информации, при этом используя модели глубокого обучения для быстрого распознавания более сложной вредоносной информации. Когда пользователь считает, что его контент не является вредоносным, он может обжаловать это, используя модель краудсорсинга пользователей для точного распознавания. Аналогично, когда пользователь сообщает о каком-то контенте, он может сначала пройти быстрое распознавание с помощью модели глубокого обучения, а затем пройти точное определение с помощью модели краудсорсинга пользователей, предложенной в этом исследовании.

2. Сравнение различных моделей краудсорсинга с различными предустановленными параметрами

Модель, созданная в этом исследовании, сравнивается с моделью краудсорсинга Zhihu. Кроме того, изменяются предустановленные параметры, чтобы создать более сложный набор данных для симуляции.

Предустановленные параметры для более сложного набора данных для симуляции следующие:

Для пользовательского набора данных, более сложный набор данных для симуляции устанавливает количество пользователей равное 100,000.

- Для `user_level` предполагается, что он следует нормальному распределению со средним значением 65 и стандартным отклонением 15. По сравнению с оригинальным набором данных для симуляции, среднее значение `user_level` уменьшается, что указывает на меньшее количество доверительных пользователей.

- Для `participation_probability` предполагается, что он следует нормальному распределению со средним значением 0.6 и стандартным отклонением 0.2. По сравнению с оригинальным набором данных для симуляции, среднее значение `participation_probability` уменьшается, что указывает на более низкую участие.

- Для `correct_probabilities` предполагается, что он следует нормальному распределению со средним значением 0.7 и стандартным отклонением 0.1. По сравнению с оригинальным набором данных для симуляции, среднее значение `participation_probability` уменьшается, что указывает на меньшую точность. Кроме того, были внесены незначительные корректировки крайних значений данных.

Предустановленные параметры для более сложного набора данных пользователя приведены в следующей таблице 6.

Таблица 6 – Предустановленные параметры для более сложного набора данных пользователя

Параметр	Среднее значение	Стандартное отклонение
user_levels	70	10
participation_probabilities	0.7	0.2
correct_probabilities	0.8	0.1
selected_times	1000	100

Для набора данных отчетов, более сложный набор данных для симуляции устанавливает количество отчетов равное 10,000.

- Для report\_types предполагается, что распределение значений 0, 1 и 2 равно соответственно 0.7, 0.2 и 0.1. Это означает, что будет больше сообщенного контента с высокой сложностью распознавания.

- Для данных с report\_types равным 0, предполагается, что difficulty\_factor следует нормальному распределению со средним значением 0.9 и стандартным отклонением 0.25, но гарантирует, что его значение находится между 0.8 и 1.

- Для данных с report\_types равным 1, предполагается, что difficulty\_factor следует нормальному распределению со средним значением 0.7 и стандартным отклонением 0.25, но гарантирует, что его значение находится между 0.6 и 0.8.

- Для данных с report\_types равным 2, предполагается, что difficulty\_factor следует нормальному распределению со средним значением 0.5 и стандартным отклонением 0.25, но гарантирует, что его значение находится между 0.4 и 0.6. Эти изменения означают, что все сообщенный контент будет более сложным для успешного распознавания.

Предустановленные параметры для более сложного набора данных отчетов приведены в следующей таблице 7.

Таблица 7 – Предустановленные параметры для более сложного набора данных отчетов

Параметр	Доля	Среднее значение	Стандартное отклонение	Ограничения
difficulty_factor (report_types = 0)	0.7	0.9	0.25	0.8 - 1
difficulty_factor (report_types = 1)	0.2	0.7	0.25	0.6 - 0.8
difficulty_factor (report_types = 2)	0.1	0.5	0.25	0.4 - 0.6

Результаты, полученные при тестировании различных моделей с наборами данных для симуляции различной сложности, представлены в таблице 4 , 8 , 9 , 10:

Таблица 8 – Результаты распознавания модели краудсорсинга Zhihu - набор данных для симуляции

Тип	Общее количество	Количество правильных распознаваний	Количество неправильных распознаваний
report_types = 0	6962	6948	14
report_types = 1	2518	2327	191
report_types = 2	520	301	219
report_types = 0,1,2	10000	9576	424



Таблица 9 – Результаты распознавания модели симуляции - более сложный набор данных для симуляции

Тип	Общее количество	Количество правильных распознаваний	Количество неправильных распознаваний
report_types = 0	7075	7023	52
report_types = 1	1952	1905	47
report_types = 2	973	898	75
report_types = 0,1,2	10000	9826	174

Таблица 10 – Результаты распознавания модели краудсорсинга Zhihu - более сложный набор данных для симуляции

Тип	Общее количество	Количество правильных распознаваний	Количество неправильных распознаваний
report_types = 0	7075	6950	125
report_types = 1	1952	1573	379
report_types = 2	973	441	532
report_types = 0,1,2	10000	8964	1036

После расчета, точность распознавания различных моделей на наборах данных для симуляции различной сложности представлена в таблице 11:

Таблица 11 – Результаты, полученные при тестировании различных моделей с наборами данных для симуляции различной сложности

	Оригинальный набор данных для симуляции	Более сложный набор данных для симуляции
Общая точность модели симуляции	99.64%	98.26%
Общая точность модели краудсорсинга Zhihu	95.76%	89.64%
Точность модели симуляции - простые типы	99.94%	99.27%
Точность модели краудсорсинга Zhihu - простые типы	99.80%	98.23%
Точность модели симуляции - сложные типы	99.52%	97.59%
Точность модели краудсорсинга Zhihu - сложные типы	92.41%	80.58%
Точность модели симуляции - очень сложные типы	96.15%	92.29%
Точность модели краудсорсинга Zhihu - очень сложные типы	57.88%	45.32%

Из приведенной выше таблицы можно видеть, что точность модели краудсорсинга пользователей, предложенной в этом исследовании, выше, чем точность модели краудсорсинга Zhihu. Даже в более сложном наборе данных для симуляции точность модели краудсорсинга пользователей, предложенной

в этом исследовании, снижается, но снижение гораздо меньше по сравнению с моделью краудсорсинга Zhihu.

## ЗАКЛЮЧЕНИЕ

В ходе выполнения магистерской диссертации была разработана модель распознавания вредоносной информации на основе краудсорсинга пользователей, и ее эффективность и точность были подтверждены с помощью симуляционных наборов данных и симуляционных моделей.

При тестировании с использованием набора данных для симуляции общая точность симуляционной модели распознавания вредоносной информации, основанной на краудсорсинге пользователей, достигла 99.64%. В частности, точность распознавания модели для простого типа вредоносной информации составила 99.94%, для сложного типа вредоносной информации – 99.52%, а для очень сложного типа вредоносной информации – 96.15%.

При тестировании с использованием более сложного набора данных для симуляции общая точность симуляционной модели распознавания вредоносной информации, основанной на краудсорсинге пользователей, достигла 98.26%. В частности, точность распознавания модели для простого типа вредоносной информации составила 99.27%, для сложного типа вредоносной информации – 97.59%, а для очень сложного типа вредоносной информации – 92.29%.

В сравнении с моделью краудсорсинга пользователей Zhihu, симуляционная модель во всех отношениях показала более высокую точность.

В рамках магистерской диссертации были решены следующие задачи:

1. Исследовались социальные сетевые платформы и вредоносная информация.
2. Изучались текущие методы распознавания вредоносной информации, вместе с их преимуществами и недостатками.
3. Создавались модель распознавания вредоносной информации на основе краудсорсинга пользователей.
4. Разрабатывались симуляционные наборы данных с помощью Python.

5. Создавались симуляционные модели с использованием Python.
6. Подтверждались эффективность и точность модели с использованием симуляционных наборов данных и симуляционных моделей.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Воронкин Алексей Сергеевич Социальные сети: эволюция, структура, анализ // ОТО. 2014. №1. URL: <https://cyberleninka.ru/article/n/sotsialnye-seti-evolyutsiya-struktura-analiz> (дата обращения: 11.05.2023).
2. Kietzmann J. H., Hermkens K., McCarthy I. P., Silvestre B. S. Social media? Get serious! Understanding the functional building blocks of social media // Business Horizons. – 2011. – Vol. 54, Issue 3. – P. 241-251.
3. Watermeyer R. Social Networking Sites // Encyclopedia of Applied Ethics (Second Edition). - Editor(s): Ruth Chadwick. - Academic Press, 2012. - P. 152-159.
4. Krasnova H., Spiekermann S., Koroleva K., et al. Online social networks: Why we disclose // Journal of information technology. - 2010. - Vol. 25, № 2. - P. 109-125.
5. Shu K., Sliva A., Wang S., et al. Fake news detection on social media: A data mining perspective // ACM SIGKDD explorations newsletter. – 2017. – Vol. 19, № 1. – P. 22-36.
6. Chu Z., Gianvecchio S., Wang H., et al. Who is tweeting on Twitter: human, bot, or cyborg? // Proceedings of the 26th annual computer security applications conference. - 2010. - P. 21-30.
7. Chu Z., Gianvecchio S., Wang H., et al. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? // IEEE Transactions on dependable and secure computing. – 2012. – Vol. 9, № 6. – P. 811-824.
8. Salahdine F., Kaabouch N. Social engineering attacks: A survey // Future Internet. – 2019. – Vol. 11, № 4. – P. 89.
9. Androutsopoulos I., Koutsias J., Chandrinou K. V., et al. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with

personal e-mail messages // Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. – 2000. – P. 160-167.

10. Ullmann S., Tomalin M. Quarantining online hate speech: technical and ethical perspectives // Ethics and Information Technology. – 2020. – Vol. 22. – P. 69-80.

11. Rawat R., Mahor V., Chirgaiya S., et al. Sentiment analysis at online social network for cyber-malicious post reviews using machine learning techniques // Computationally intelligent systems and their applications. – 2021. – P. 113-130.

12. Kerr A., Kelleher J. D. The recruitment of passion and community in the service of capital: Community managers in the digital games industry // Critical studies in media communication. – 2015. – Vol. 32, № 3. – P. 177-192.

13. Myers West S. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms // New Media & Society. – 2018. – Vol. 20, № 11. – P. 4366-4383.

14. de Saint Laurent C., Glaveanu V., Chaudet C. Malevolent creativity and social media: Creating anti-immigration communities on Twitter // Creativity Research Journal. – 2020. – Vol. 32, № 1. – P. 66-80.

15. Deibert R. J. The road to digital unfreedom: Three painful truths about social media // Journal of Democracy. – 2019. – Vol. 30, № 1. – P. 25-39.

16. Myers West S. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms // New Media & Society. – 2018. – Vol. 20, № 11. – P. 4366-4383.

17. Gill Z. User-driven collaborative intelligence: social networks as crowdsourcing ecosystems // CHI'12 Extended Abstracts on Human Factors in Computing Systems. – 2012. – P. 161-170.

18. Zhihu Crowdsourcing Terms (Trial) - Zhihu - URL:  
<https://www.zhihu.com/court/terms> (дата обращения: 15. 05. 2023).



# Отчет о проверке на заимствования №1



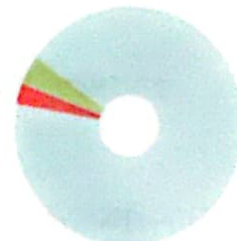
Автор: Чжу Сюэфэн  
Проверяющий: Романович Ольга Владимировна  
Организация: Томский Государственный Университет  
Отчет предоставлен сервисом «Антиплагиат» - <http://tsu.antiplagiat.ru>

## ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 18  
Начало загрузки: 04.06.2023 08:27:43  
Длительность загрузки: 00:00:12  
Имя исходного файла: Модель  
распознавания вредоносной информации  
на основе краудсорсинга  
пользователей.docx  
Название документа: Модель  
распознавания вредоносной информации  
на основе краудсорсинга пользователей  
Размер текста: 99 кБ  
Символов в тексте: 101525  
Слов в тексте: 11542  
Число предложений: 671

## ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Начало проверки: 04.06.2023 08:27:56  
Длительность проверки: 00:01:12  
Корректировка от 04.06.2023 08:30:28  
Комментарии: [Автосохраненная версия]  
Поиск с учетом редактирования: да  
Проверенные разделы: титульный лист с. 1, основная часть с. 2-4,6-73,  
содержание с. 5, библиография с. 74-76  
Модули поиска: ИПС Адилет, Библиография, Сводная коллекция ЭБС, Интернет  
Плюс\*, Сводная коллекция РГБ, Цитирование, Переводные заимствования (RuEn),  
Переводные заимствования по eLIBRARY.RU (EnRu), Переводные заимствования  
по коллекции Гарант: аналитика, Переводные заимствования по коллекции  
Интернет в английском сегменте, Переводные заимствования по Интернету  
(EnRu), Переводные заимствования по коллекции Интернет в русском сегменте,  
Переводные заимствования издательства Wiley, eLIBRARY.RU, СПС ГАРАНТ:  
аналитика, СПС ГАРАНТ: нормативно-правовая документация, Медицина,  
Диссертации НББ, Коллекция НБУ, Перефразирования по eLIBRARY.RU,  
Перефразирования по СПС ГАРАНТ: аналитика, Перефразирования по Интернету,  
Перефразирования по Интернету (EN), Перефразированные заимствования по  
коллекции Интернет в английском сегменте, Перефразированные заимствования  
по коллекции Интернет в русском сегменте, Перефразирования по коллекции  
издательства Wiley, Патенты СССР, РФ, СНГ, СМИ России и СНГ, Шаблонные  
фразы, Модуль поиска "tsu", Кольцо вузов, Издательство Wiley, Переводные  
заимствования



### СОВПАДЕНИЯ

3,29%

### САМОЦИТИРОВАНИЯ

0%

### ЦИТИРОВАНИЯ

4,41%

### ОРИГИНАЛЬНОСТЬ

92,3%

**Совпадения** — фрагменты проверяемого текста, полностью или частично сходные с найденными источниками, за исключением фрагментов, которые система отнесла к цитированию или самоцитированию. Показатель «Совпадения» — это доля фрагментов проверяемого текста, отнесенных к совпадениям, в общем объеме текста.

**Самоцитирования** — фрагменты проверяемого текста, совпадающие или почти совпадающие с фрагментом текста источника, автором или соавтором которого является автор проверяемого документа. Показатель «Самоцитирования» — это доля фрагментов текста, отнесенных к самоцитированию, в общем объеме текста.

**Цитирования** — фрагменты проверяемого текста, которые не являются авторскими, но которые система отнесла к корректно оформленным. К цитированиям относятся также шаблонные фразы; библиография; фрагменты текста, найденные модулем поиска «СПС Гарант: нормативно-правовая документация». Показатель «Цитирования» — это доля фрагментов проверяемого текста, отнесенных к цитированию, в общем объеме текста.

**Текстовое пересечение** — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.

**Источник** — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.

**Оригинальный текст** — фрагменты проверяемого текста, не обнаруженные ни в одном источнике и не отмеченные ни одним из модулей поиска. Показатель «Оригинальность» — это доля фрагментов проверяемого текста, отнесенных к оригинальному тексту, в общем объеме текста.

«Совпадения», «Цитирования», «Самоцитирования», «Оригинальность» являются отдельными показателями, отображаются в процентах и в сумме дают 100%, что соответствует полному тексту проверяемого документа.

Обращаем Ваше внимание, что система находит текстовые совпадения проверяемого документа с проиндексированными в системе источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности совпадений или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

№	Доля в тексте	Доля в отчете	Источник	Актуален на	Модуль поиска	Комментарии
[01]	3,41%	3,41%	не указано	29 Сен 2022	Библиография	
[02]	1,57%	0,91%	диплом.pdf	02 Июнь 2022	Модуль поиска "tsu"	
[03]	1,48%	0,5%	VKR_Subach-2.docx	16 Янв 2022	Модуль поиска "tsu"	
[04]	1,38%	0%	ВКР Субач-2.docx	16 Янв 2022	Модуль поиска "tsu"	
[05]	1,36%	0,1%	СТРУКТУРА ЛАНДШАФТОВ И КАЧЕСТВЕННАЯ ОЦЕНК...	09 Июнь 2022	Модуль поиска "tsu"	
[06]	1,21%	0,17%	Силур-девонские кораллы Горного Алтая из коллек...	05 Июнь 2022	Модуль поиска "tsu"	
[07]	1,19%	1,19%	53758 <a href="http://e.lanbook.com">http://e.lanbook.com</a>	09 Мар 2016	Сводная коллекция ЭБС	
[08]	1,01%	1%	не указано	29 Сен 2022	Шаблонные фразы	

*С результатами согласованно*