МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение  
высшего образования

**«Южно-Уральский государственный университет**

**(национальный исследовательский университет)»**

**Высшая школа электроники и компьютерных наук**

**Кафедра системного программирования**

|  |
| --- |
| ДОПУСТИТЬ К ЗАЩИТЕ  Заведующий кафедрой, д.ф.-м.н., профессор  \_\_\_\_\_\_\_\_\_\_ Л.Б. Соколинский  «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_ 2023 г. |

**Реализация системы на основе методов машинного анализа географически-распределенных данных**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ЮУрГУ – 09.03.04.2023.308-278.ВКР

|  |  |
| --- | --- |
|  | Научный руководитель,  доцент кафедры СП, к.ф.-м.н. \_\_\_\_\_\_\_\_\_\_ Г.И. Радченко  Автор работы, студент группы КЭ-403  \_\_\_\_\_\_\_\_\_\_ В.А. Дегтярев  Ученый секретарь  (нормоконтролер)  \_\_\_\_\_\_\_\_\_\_\_\_\_ И.Д. Володченко  «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_ 2023 г. |

Челябинск, 2023 г.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение  
высшего образования

**«Южно-Уральский государственный университет**

**(национальный исследовательский университет)»**

**Высшая школа электроники и компьютерных наук**

**Кафедра системного программирования**

УТВЕРЖДАЮ

Зав. кафедрой СП

\_\_\_\_\_\_\_\_\_\_ Л.Б. Соколинский

06.02.2023 г.

**ЗАДАНИЕ**

**на выполнение выпускной квалификационной работы бакалавра**

студенту группы КЭ-403

Дегтяреву Владимиру Андреевичу,

обучающемуся по направлению

09.03.04 «Программная инженерия»

1. **Тема работы** (утверждена приказом ректора от \_\_.\_\_.2022 г. № \_\_\_)

Реализация системы на основе методов машинного анализа географически‑распределенных данных.

1. **Срок сдачи студентом законченной работы:** 05.06.2023 г.
2. **Исходные данные к работе2**
3. Официальный сайт OpenMined. [Электронный ресурс] URL: <https://www.openmined.org/> (дата обращения: 13.02.2023 г.).
4. Официальный сайт PySyft. [Электронный ресурс] URL: <https://github.com/OpenMined/PySyft> (дата обращения 13.02.2023 г.).
5. Официальный сайт TenserFlow Federated. [Электронный ресурс] URL: <https://www.tensorflow.org/federated?hl=ru> (дата обращения 13.02.2023 г.).
6. **Перечень подлежащих разработке вопросов**
7. Выполнить обзор литературы.
8. Выполнить анализ аналогичных проектов.
9. Определить функциональные и нефункциональные требования к системе.
10. Спроектировать методы машинного анализа географически‑распределенных данных.
11. Реализовать методы машинного анализа географически‑распределенных данных.
12. Провести тестирование методов машинного анализа географически‑распределенных данных.
13. **Дата выдачи задания:** 06.02.2023 г.

**Научный руководитель,**

доцент кафедры СП, к.ф.-м.н. Г.И. Радченко

**Задание принял к исполнению** В.А. Дегтярев

ОГЛАВЛЕНИЕ

[ВВЕДЕНИЕ 5](#_Toc127213919)

[1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ 7](#_Toc127213920)

[1.1. Описание предметной области 7](#_Toc127213921)

[1.1. Анализ аналогичных проектов 7](#_Toc127213922)

[ЗАКЛЮЧЕНИЕ 10](#_Toc127213923)

[ЛИТЕРАТУРА 11](#_Toc127213924)

# ВВЕДЕНИЕ

**Актуальность**

В настоящее время в мире существует огромное количество информации. А также большое количество людей, которые хотят использовать эту информацию в своих научных интересах, для нахождения закономерностей, обучения нейросетей, чтобы получать ответы на важные вопросы.

Машинное обучение и анализ данных уже активно применяется в медицине, финансах, промышленности [1]. Однако эти технологии еще не могут уверенно отвечать на некоторые глобальные и сложные вопросы из-за отсутствия доступа у разработчиков и ученых к большому количеству информации. Основные причины этого выражаются в виде защиты персональных данных, сохранения приватности конкретных данных, а также раздробленности этих данных среди огромного количества организаций.

Эти проблемы можно решить с помощью систем географически-распределенного и конфиденциального машинного обучения.

**Постановка задачи**

Целью выпускной квалификационной работы является реализация системы на основе методов машинного анализа географически‑распределенных данных. Для достижения поставленной цели необходимо решить следующие задачи:

1. выполнить обзор литературы;
2. выполнить анализ аналогичных проектов;
3. определить функциональные и нефункциональные требования к системе;
4. спроектировать методы машинного анализа географически‑распределенных данных;
5. реализовать методы машинного анализа географически‑распределенных данных;
6. провести тестирование методов машинного анализа географически‑распределенных данных.

**Структура и содержание работы**

Работа состоит из введения, пяти глав, заключения и списка литературы. Объем работы составляет 29 страниц, объем списка литературы – 11 источников.

В первой главе описываются предметная область и аналогичные проекты.

Вторая глава содержит описание теоретической части по теме работы.

Третья глава посвящена определению функциональных и нефункциональных требований к системе и проектированию ее архитектуры.

Четвертая глава содержит в себе подробности и особенности реализации методов машинного анализа.

В пятой главе описывается процесс тестирования работы методов машинного анализа географически-распределенных данных.

# 1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

1.1. Описание предметной области

Целью данной работы является разработка системы на основе методов машинного анализа географически-распределенных данных. Анализ данных представляет собой область математики и информатики, которая занимается построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных данных с целью получения полезной информации и принятия решений [2].

Особенностью машинного анализа географически-распределенных данных является то, что информация, которую анализирует система, распределена между различными независимыми устройствами.

1.1. Анализ аналогичных проектов

**PySyft**

В 2019 году была создана библиотека PySyft сообществом OpenMined. Это люди, объединенные темой конфиденциальности в машинном обучении. PySyft представляет собой обертку над PyTorch, Tensorflow или Keras для приватного машинного обучения [1].

Основная задача, стоящая перед сообществом OpenMined, заключалась в том, чтобы создать программное обеспечение, которое бы позволяла одному человеку получать ответы на свои вопросы, используя данные, принадлежащие другому человеку без необходимости просмотра и создания копии этих данных [2].

Проект предоставляет удаленный вызов процедур, что позволяет разработчику отправлять нейросеть к пользователям, где она локально обучается на их данных, после чего возвращается с обновленными весами обратно разработчику (рисунок 1). Данный процесс может происходить одновременно на разных устройствах, тем самым происходит параллельное обучение нейросети [1].



Рисунок 1 – Взаимодействие разработчика с пользователями [ссылка]

Также ключевой особенностью PySyft является использование дифференциальной приватности [2]. По измененным весам нейросети можно догадаться, какие данные были у пользователя. Чтобы это предотвратить, пользователь может добавить в данные шум. Дифференциальная приватность представляет собой методы, которые описывают добавление шума.

**TensorFlow Federated**

Компания Google давно занимается сбором некоторой информации с устройств пользователей в единое защищенное хранилище, на котором тренируют свои нейросети. А в 2017 году ученые из Google Research предложили инновационный подход под названием федеративное машинное обучение. Он позволяет всем устройствам, которые участвуют в машинном обучении, делить на всех единую модель для прогнозирования, но при этом не делиться первичными данными для обучения модели. Система федеративного обучения работает по принципу совершенствования единой общей модели нейросети [5].

Компанией Google была создана платформа TensorFlow Federated с открытым исходным кодом для машинного обучения на децентрализованных данных. Архитектура данного платформы представлена на рисунке 2. TensorFlow Federated был разработан для облегчения исследований и экспериментов с федеративным обучением [6].



Рисунок 2 – Архитектура TensorFlow Federated

**Выводы по первой главе**

В данной главе был проведен обзор предметной области и анализ существующих решений и проектов в области машинного анализа географически-распределенных данных.

# ЗАКЛЮЧЕНИЕ

В рамках данной работы был реализована система на основе методов машинного анализа географически‑распределенных данных. При этом были решены следующие задачи.

1. Выполнить обзор литературы.
2. Выполнить анализ аналогичных проектов.
3. Определить функциональные и нефункциональные требования к системе.
4. Спроектировать методы машинного анализа географически‑распределенных данных.
5. Реализовать методы машинного анализа географически‑распределенных данных.
6. Провести тестирование методов машинного анализа географически‑распределенных данных.

# ЛИТЕРАТУРА

1. 7 примеров применения машинного обучения в 5 отраслях бизнеса. [Электронный ресурс] URL: https://mcs.mail.ru/blog/17-primerov-mashinnogo-obucheniya (дата обращения: 13.02.2023 г.).
2. Анализ данных – основы и терминология. [Электронный ресурс] URL: https://habr.com/ru/post/352812/ (дата обращения: 13.02.2023 г.).
3. Конфиденциальное машинное обучение. Библиотека PySyft. [Электронный ресурс] URL: <https://habr.com/ru/post/500154/> (дата обращения: 13.02.2023 г.).
4. Официальный сайт OpenMined. [Электронный ресурс] URL: <https://www.openmined.org/> (дата обращения: 13.02.2023 г.).
5. Google изобрела распределённый ИИ для миллиарда смартфонов. [Электронный ресурс] URL: https://habr.com/ru/post/402987/ (дата обращения: 13.02.2023 г.).
6. Официальный сайт Tensorflow Federated. [Электронный ресурс] URL: https://www.tensorflow.org/federated/ (дата обращения: 13.02.2023 г.).