Федеральное государственное автономное

образовательное учреждение высшего образования

«Пермский государственный национальный исследовательский университет» (ПГНИУ)

Региональный институт непрерывного образования (РИНО ПГНИУ)

Цифровая кафедра

Выпускная аттестационная (квалификационная) работа

по курсу профессиональной переподготовки

«Анализ данных и машинное обучение»

**АНАЛИЗ ТОНАЛЬНОСТИ ОТЗЫВОВ**

Разработчики проекта:

Азматова Карина Василевна

Кравчук Юлия Сергеевна

Шмырина Владислава Олеговна

Пермь, 2024

**ПАСПОРТ ПРОЕКТА**

**Название проекта:**

Анализ тональности отзывов

**Сведения об авторах:**

Азматова Карина Василевна

Кравчук Юлия Сергеевна

Шмырина Владислава Олеговна

**Цель и задачи проекта:**

**Цель:**

Разработка и обучение моделей машинного обучения и глубокого обучения для анализа тональности отзывов, с целью определения их эмоциональной окраски (положительная, отрицательная или нейтральная).

**Задачи:**

 Изучить предоставленный набор данных:

* Проверить наличие пропусков и аномалий (выбросов).
* Исследовать распределение классов для оценки балансировки.
* Провести предварительную обработку данных, включая очистку текста, токенизацию и векторизацию.

 Выбрать три модели для решения задачи:

* Минимум одну модель из классического машинного обучения.
* Минимум одну модель на основе глубокого обучения.

 Определить функцию потерь для задачи анализа тональности.

 Выбрать метрики оценки качества моделей.

 Произвести обучение и тестирование моделей.

 Сравнить качество моделей и выбрать наилучшую для решения задачи.

**Краткое описание проекта:**

Проект посвящён автоматическому анализу тональности отзывов на основе текста и оценок, полученных с использованием веб-скрапинга. Набор данных содержит текстовые отзывы и их метаинформацию, включая оценку пользователя. На основе обработанного набора данных будут обучены модели машинного и глубокого обучения для классификации отзывов.

**Конкретные ожидаемые результаты:**

 Набор данных, готовый для обучения моделей.

 Три обученные модели: минимум одна из классического машинного обучения и минимум одна нейросетевая модель.

 Наглядная демонстрация результатов работы моделей для сравнения с выводом наилучшей из них.