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»

**Классификация последовательности DNA Seq**

Разработчики проекта:

Студенты группы ПМИ-1,2-2021

Андрюков Анатолий Иванович,

Ковбаснюк Владимир Васильевич,

Лис-Граундер Алиса Васильевна

Пермь, 2024

**ПАСПОРТ ПРОЕКТА**

**Название проекта:**

Классификация последовательности DNA Seq

**Сведения об авторах:**

Андрюков А.И., Ковбаснюк В.В., Лис-Граундер А.В.

**Цель и задачи проекта:**

**Цель:**

На основе данных о последовательностях ДНК человека и их классов создать три модели классификации последовательностей ДНК и выбрать лучшую из них.

**Задачи:**

1. Изучить набор данных;
2. Выбрать три модели классификации;
3. Выбрать критерии качества решения задачи;
4. Произвести тренировку моделей
5. Выбрать лучшую модель.

**Краткое описание проекта:**

Изучить и проанализировать данные о последовательностях ДНК человека, представленных в виде символьной последовательности нуклеотидов. Построить классификаторы для последовательностей ДНК и оценить их работу.

**Конкретные ожидаемые результаты:**

Создать три классификатора последовательностей ДНК человека и оценка их работы. После выбора конкретных моделей (KNN, дерево решений, модель глубокого обучения) ожидается, что лучшей моделью окажется KNN.