

ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий

Кафедра «Информатика»

Технологии анализа данных

Практическая работа № 2

Инженерия машинного обучения

Красноярск, 2022

Цель: знакомство с инструментами организации процесса MLOps, получение практических навыков по работе с версионированием экспериментов и автоматизации пайплайнов машинного обучения.

Задачи

Выполнение практической работы предполагает решение следующих задач.

1. Создать удаленный репозиторий для проекта.
2. Разработать решение задачи в соответствии с вариантом.
 - a. Построить автоматический ML-пайплайн, состоящий из всех необходимых шагов для решения задачи (предобработка данных, обучение моделей, оценка качества и т.д.).
 - b. Провести серию экспериментов, используя как минимум три различные модели машинного обучения для решения поставленной задачи. При защите работы необходимо продемонстрировать возможность перемещения по истории экспериментов с целью выбора конкретной модели.
3. Выполнить аугментацию, сгенерировав дополнительные данные с помощью различных преобразований, и сохранить новый датасет.
4. Повторить пункт 2, используя все данные (исходные и аугментированные).
5. В отчете привести сравнение результатов, получаемых при различных условиях экспериментов на разных наборах данных (с аугментацией и без).

Варианты

Базовые варианты заданий приведены в списке далее. Приветствуется использование собственной задачи.

- 1) Butterfly Specie Classification Dataset
- 2) Food vs Non-Food Image Dataset
- 3) Zalando Store Fashion Image Dataset

4) Plant Pathology FGVC7 Dataset

Материально-техническое обеспечение:

Студенту предлагается использовать GitLab для создания удаленного репозитория и DVC (<https://dvc.org/>) для организации пайплайна и экспериментов. Допускается использование других инструментов (рассмотренных в лекции или найденных самостоятельно студентом), позволяющих выполнить все поставленные задачи.

Требования к выполнению практической работы:

1. Написание программного кода и формирование результатов согласно заданию и установленному варианту.
2. Составление отчета, содержащего описание решаемых задач методов решения и полученных результатов.
3. Воспроизводимость полученного результата.