Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Программной инженерии и компьютерной техники

**Отчёт**

**Лабораторная работа №4**

**Освоение инструментов для работы с гипотезами и экспериментами в программной системе ИИ.**

**Ф.И.О. студента** Герасимчук Михаил Юрьевич

**Группа** P4141

**Преподаватель** Старобыховская А.А.

***г. Санкт-Петербург, 2022***

**Оглавление**

[Задание 3](#_Toc123087255)

[Цель: 3](#_Toc123087256)

[Описание предметной области: 3](#_Toc123087257)

[Исходные данные: 3](#_Toc123087258)

[Ход выполнения работы 3](#_Toc123087259)

[Вывод 5](#_Toc123087260)

# Задание

## Цель:

Настройка системы версионирования данных и гипотез DVC.

## Описание предметной области:

Эксперименты при создании программной системы ИИ.

## Исходные данные:

Открытый набор данных по ссылке - http:// archive.ics.uci.edu/ml/datasets/Iris.

# Ход выполнения работы

Во время выполнения был создан git репозиторий, который был загружен на Gitlab. Окружение Python получается с использованием сгенеренного venv файла. Необходимые пакеты описаны в файле requirements.txt, при запуске которого можно получить рабочее окружение для работы с DVC. Ниже представлена итоговая версия созданных веток. Этапы выполнения задания можно отследить через коммиты в Gitlab.

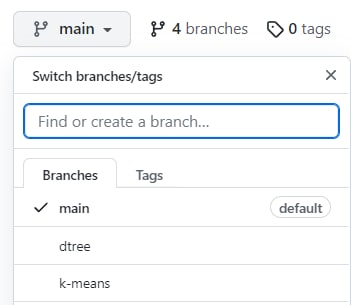


Рис. 1 Ветки созданные в ходе работы

DAG работает следующим образом:

* выполняется файл prepare.py, разделяющий набор данных на тренировочную и тестовые выборки, отделение целевой переменной и предобработка данных.

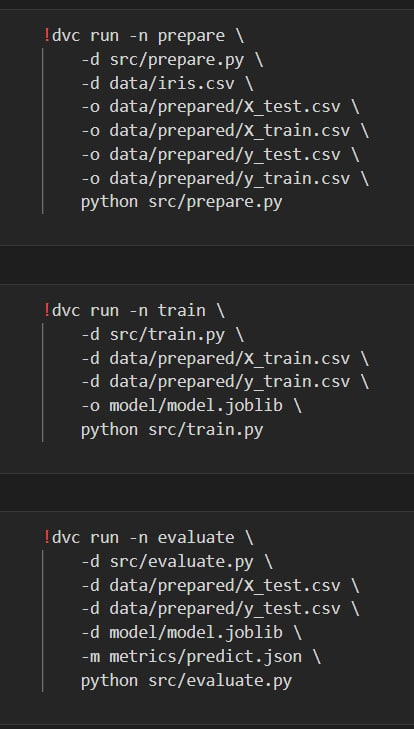


Рис. 2 Создание DAG

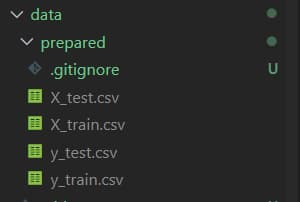


Рис. 3 Тестовая и тренировочная выборки

Следующий этап это обучение модели на сгенерированных данных, модель сохраняется в файл model.joblib. После чего наступает этап расчёта метрики accuracy, получаемый с помощью тестового набора. Результаты записываются в файл predict.json.



Рис. 4 Результат проверки на тестовом наборе

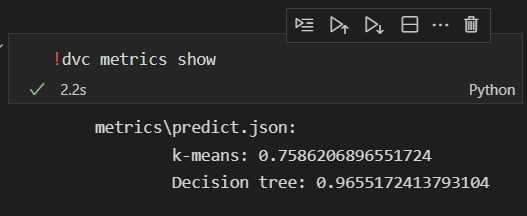


Рис. 5 Результаты экспериментов

# Вывод

В процессе выполнения данной лабораторной работы мы получили опыт работы с DVC. Научились автоматизировать процесс эксперимента с использованием DAG, применили различные модели машинного обучения k-means, Decision tree и оценили точность их работы.