

Оформляем единый github-репозиторий, в котором представлены результаты всех 5 лабораторных работ. Github-репозиторий должен обязательно содержать readme + 5 jupyter-notebook'ов, в которых даны комментарии как минимум к каждой ячейке с кодом.

Лабораторная работа №1 (Проведение исследований с алгоритмом KNN)

1. Выбор начальных условий

- a. Выбрать набор данных для задачи классификации (у каждого студента должен быть уникальный набор данных) и обосновать его выбор (реальная практическая задача)
- b. Выбрать набор данных для задачи регрессии (у каждого студента должен быть уникальный набор данных) и обосновать его выбор (реальная практическая задача)
- c. Выбрать метрики качества и обосновать их выбор

2. Создание бейзлайна и оценка качества

- a. Обучить модели из sklearn (для классификации и регрессии) для выбранных наборов данных
- b. Оценить качество моделей (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных

3. Улучшение бейзлайна

- a. Сформулировать гипотезы (препроцессинг данных, визуализация данных, формирование новых признаков, подбор гиперпараметров на кросс-валидации и т.д.)
- b. Проверить гипотезы
- c. Сформировать улучшенный бейзлайн по результатам проверки гипотез
- d. Обучить модели с улучшенным бейзлайном (для классификации и регрессии) для выбранных наборов данных

- e. Оценить качество моделей с улучшенным бейзлайном (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
- f. Сравнить результаты моделей с улучшенным бейзлайном в сравнении с результатами из пункта 2
- g. Сделать выводы

4. Имплементация алгоритма машинного обучения

- a. Самостоятельно имплементировать алгоритмы машинного обучения (для классификации и регрессии)
- b. Обучить имплементированные модели (для классификации и регрессии) для выбранных наборов данных
- c. Оценить качество имплементированных моделей (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
- d. Сравнить результаты имплементированных моделей в сравнении с результатами из пункта 2
- e. Сделать выводы
- f. Добавить техники из улучшенного бейзлайна (пункт 3с)
- g. Обучить модели (для классификации и регрессии) для выбранных наборов данных
- h. Оценить качество моделей (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
- i. Сравнить результаты моделей в сравнении с результатами из пункта 3
- j. Сделать выводы

Лабораторная работа №2 (Проведение исследований с логистической и линейной регрессией) – повтор пунктов 2-4 из лабораторной работы №1

Лабораторная работа №3 (Проведение исследований с решающим деревом) – повтор пунктов 2-4 из лабораторной работы №1

Лабораторная работа №4 (Проведение исследований со случайным лесом) – повтор пунктов 2-4 из лабораторной работы №1

Лабораторная работа №5 (Проведение исследований с градиентным бустингом) – повтор пунктов 2-4 из лабораторной работы №1 + подведение итогов / сравнение результатов, полученных всеми алгоритмами из лабораторных 1-5)