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

Лабораторная работа №1 (Проведение исследований с алгоритмом KNN)

1. Выбор начальных условий
   1. Выбрать набор данных для задачи классификации (у каждого студента должен быть уникальный набор данных) и обосновать его выбор (реальная практическая задача)
   2. Выбрать набор данных для задачи регрессии классификации (у каждого студента должен быть уникальный набор данных) и обосновать его выбор (реальная практическая задача)
   3. Выбрать метрики качества и обосновать их выбор
2. Создание бейзлайна и оценка качества
   1. Обучить модели из sklearn (для классификации и регрессии) для выбранных наборов данных
   2. Оценить качество моделей (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
3. Улучшение бейзлайна
   1. Сформулировать гипотезы (препроцессинг данных, визуализация данных, формирование новых признаков, подбор гиперпараметров на кросс-валидации и т.д.)
   2. Проверить гипотезы
   3. Сформировать улучшенный бейзлайн по результатам проверки гипотез
   4. Обучить модели с улучшенным бейзлайном (для классификации и регрессии) для выбранных наборов данных
   5. Оценить качество моделей с улучшенным бейзлайном (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
   6. Сравнить результаты моделей с улучшенным бейзлайном в сравнении с результатами из пункта 2
   7. Сделать выводы
4. Имплементация алгоритма машинного обучения
   1. Самостоятельно имплементировать алгоритмы машинного обучения (для классификации и регрессии)
   2. Обучить имплементированные модели (для классификации и регрессии) для выбранных наборов данных
   3. Оценить качество имплементированных моделей (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
   4. Сравнить результаты имплементированных моделей в сравнении с результатами из пункта 2
   5. Сделать выводы
   6. Добавить техники из улучшенного бейзлайна (пункт 3с)
   7. Обучить модели (для классификации и регрессии) для выбранных наборов данных
   8. Оценить качество моделей (для классификации и регрессии) по выбранным метрикам на выбранных наборах данных
   9. Сравнить результаты моделей в сравнении с результатами из пункта 3
   10. 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)

Таблица 1 Метрика качества на тестовом наборе данных

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Алгоритм** | **Задача** | **Бейзлайн** | **Улучшенный бейзлайн** | **Самостоятельная имплементация алгоритма** |
| KNN | классификация |  |  |  |
| регрессия |  |  |  |
| Линейные модели | классификация |  |  |  |
| регрессия |  |  |  |
| Решающее дерево | классификация |  |  |  |
| регрессия |  |  |  |
| Случайный лес | классификация |  |  |  |
| регрессия |  |  |  |
| Градиентный бустинг | классификация |  |  |  |
| регрессия |  |  |  |