Лабораторная 2

Лабораторная посвящена закреплению пройденного материала по методам классического машинного обучения.Лабораторная обязательно сохраняется в формате .ipynb(т.е. код разбит на ячейки и сохранен вывод каждой из них)

Для выполнения лабораторной необходимо:

1. Скачать или импортировать в **notebook** датасет [Wine](https://archive.ics.uci.edu/dataset/109/wine), после чего отобразить первые и последние 5 строк датасета
2. Необходимо рассмотреть данные и провести анализ( количество семплов, типы данных в ячейках, распределение, корреляция…). Желательно придти к какому-то выводу после рассмотрения
3. Произвести визуализации пункта 2, а также провести анализ с помощью библиотек matplotlib,seaborn
4. Разбить данные на test и train. Предварительно убрав колонки с таргетом
5. Обучить decision tree на 2х самых коррелирующих признаках и не коррелирующих признаках, используя gini. Посмотреть метрики(f1 score,accuracy,precision,recall), сделать выводы, построить ConfusionMatrix
6. Обучить на выбор (SVM, XGboost,RandomForest), на всех признаках. Получить результат, отобразить метрики как в пункте 5, в случае обучения методов использующих деревья отразить важность признаков
7. Построить график roc curve

Критерии:

* Работа сдана в формате notebook (2б)
* Пункт 1 (3б)
* Пункты 2,3,4 (2б+2б+1б)
* Пункт 5 (5б)
* Пункты 6-7(3б+2б)
* Баллы будут зачтены при прохождении процедуры защиты

Защита:

Для защиты необходимо, чтобы notebook находился в репозитории в отдельной папке с названием lab\_2. После загрузки решения необходимо договориться о времени для созвона. Чтобы работа была зачтена, необходимо представить свое решение, после чего ответить на вопросы преподавателя, если они возникнут. При ответе на все вопросы, выставляются баллы. В случае ошибок в работе(или вопросов на которые не было дано ответа), есть возможность исправить их до дедлайна, после чего заново пройти процедуру защиты.

**DEADLINE 24.02.24**