**Лабораторная работа № 5**

**Тема**: Бинарная классификация.

**Цель работы** *–* изучить на практике понятие бинарной классификации и методы ее реализации из библиотеки *sklearn*: деревья решений и метод k-ближайших соседей*.*

1. Выберите любой доступный на просторах интернета набор данных (dataset) подходящий для бинарной классификации. Например, среди предложенных наборов:

<https://www.kaggle.com/datasets?search=binary+classification>

!!! В подгруппе датасеты должны быть уникальными, работа с повторяющимся датасетом не принимается. Сумейте договориться друг с другом. В крайнем случае работы должны быть уникальными с авторскими рассуждениями и решениями.

Требования к набору данных:

- минимум пять признаков, описывающих объект (можно больше);

- минимум 100 объектов для обучающего набора (можно больше);

- по набору можно оценить вероятность какого-то бинарного исхода (например, выживет/не выживет, здоровый/больной, женат/неженат, кредит выдавать/не выдавать и т.п.)

2. Проанализируйте исходные данные, при необходимости заполните пропуски или удалить не важную информацию. Категориальные признаки замените на числовые

3. Выделите из данных вектор меток У и матрицу признаков Х.

4. Разделите набор данных на обучающую и тестовую выборки

5. На обучающей выборке получите модели дерева решений и k-ближайших соседей, используйте кросс валидацию.

6. Подберите наилучшие гиперпараметры моделей (например, глубину для дерева решений, количество соседей для алгоритма knn)

7. Выберите лучшую модель, оцените ее качество.

8. Визуализируйте полученную модель (при визуализации дерева решений желательно уменьшить глубину дерева, что бы рисунок был читаемым, или сохранить в отдельный файл)

9. Сделать выводы и оформить электронный отчёт по результатам работы.