**Технологии прикладного анализа данных**

**Лабораторная работа №2**

**Деревья решений**

Выполните следующие задания:

1. Вы работаете с файлом winequality-red.csv. Целевая переменная – quality.
2. Выполните предобработку данных при необходимости.
3. Выполните машинное обучение по модели деревьев решений.
4. Исследуйте влияние глубины дерева на модель, используя 5-fold кросс-валидацию в диапазоне от 1 до 15. Особое внимание уделите подбору метрики. Примеры метрик приведены ниже: **Valid options are ['accuracy', 'adjusted\_mutual\_info\_score', 'adjusted\_rand\_score', 'average\_precision', 'completeness\_score', 'explained\_variance', 'f1', 'f1\_macro', 'f1\_micro', 'f1\_samples', 'f1\_weighted', 'fowlkes\_mallows\_score', 'homogeneity\_score', 'mutual\_info\_score', 'neg\_log\_loss', 'neg\_mean\_absolute\_error', 'neg\_mean\_squared\_error', 'neg\_mean\_squared\_log\_error', 'neg\_median\_absolute\_error', 'normalized\_mutual\_info\_score', 'precision', 'precision\_macro', 'precision\_micro', 'precision\_samples', 'precision\_weighted', 'r2', 'recall', 'recall\_macro', 'recall\_micro', 'recall\_samples', 'recall\_weighted', 'roc\_auc', 'v\_measure\_score']**
5. Постройте графики сравнения результатов на обучающем и тестовом множестве с указанием метрики. Необходимо подобрать хотя бы две различные метрики.
6. Выполните подбор гиперпараметров по методу случайного перебора. В пространство поиска включите следующие параметры: критерий, максимальная глубина, минимальное число объектов в листе, вес класса и минимальное количество объектов, необходимое для разделения внутреннего узла. Количество запусков не менее 200.
7. Выведите параметры лучшей модели, наилучшее значение метрики и важность признаков.

Кратко опишите полученные результаты (ко всей работе).