**Лабораторная работа № 4** (4 часа)

**Тема**: Бинарная классификация.

**Цель работы** *–* изучить на практике понятие бинарной классификации и методы ее реализации из библиотеки *sklearn.*

Примеры к этой работе можно найти в папке Examples for 4 lab.

1. Используйте датасет из предыдущих работ или выберите любой датасет, доступный на просторах интернета и **подходящий для бинарной классификации.**

2. Если вы используете новый датасет, то предобработайте данные (заполните пропуски, удалить выбросы и 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 **обучите модели** следующими методами классификации: дерево решений, случайных лес, k-ближайших соседей, SVC)

6. **Оцените точность** каждой модели, рассчитайте матрицу ошибок и метрики качества (Accuracy, Precision, Recall) постройте ROC-кривые для всех моделей в одних осях для сравнения).

7. **Улучшите модели**, например, подберите наилучшие гиперпараметры моделей при помощи функции **GridSearchCV()** (например, *max\_depth* и *max\_features* для дерева решений; *n\_neighbors* для алгоритма knn; *С и gamma* для SVC).

8. **Оцените** **точность** моделей после их улучшения, сравните с точностью до улучшения.

9. Сделать выводы о лучшей модели для данного датасета и **оформить электронный отчёт** по результатам работы (*если работаете в jupyter notebook или в каком-либо другом notebook, то можно просто сопроводить код комментариями и заголовками*).