# **Лабораторная работа #3**

**Цель работы:** построить модель регрессии на основе модели k ближайших соседей и модели классификации на основе алгоритма логистической регрессии.

**Инструменты:** модели регрессии (KNeighborsRegressor) и классификации (LogisticRegression) из пакета scikit-learn.

**Содержание:**

Для предложенных датасетов (один для классификации, другой для регрессии) построить соответствующие модели. Для этого:

1. Воспользоваться результатами Лабораторной работы #1 по препроцессингу данных.
2. Разбить данные на обучающую и тестовую выборки (например, в соотношении 80/20).
3. Построить регрессор с различными вариантами архитектур.
4. Построить классификатор при одном наборе параметров (выбрать самостоятельно).
5. Обучить модели на обучающих частях выборки.
6. Провести сравнительную оценку качества моделей на основе тестовой выборки. Для классификатора провести более подробный анализ.

## **Архитектура**

**Замечание про архитектуры (регрессор):**

1. Количество соседей ***n\_neighbors***: не менее 3 различных значе- ний. Обязательно рассмотреть случай ***n\_neighbors*** = 1;
2. Вес объектов ***weights***: рассмотреть ’*uniform*’ и ’*distance*’;
3. Метрика: метрика городских кварталов (‘*cityblock*’), евклидова метрика (‘*euclidean*’) и метрика Чебышева (*scipy.spatial.distance.chebyshev*).
4. Используя *GridSearchCV* определить лучший регрессор.

**Оценка качества классификации:**

Логистическая регрессия — это модель, которая возвращает вероятность принадлежность классу. В связи с этим предлагается

1. Построить ROC-кривую по результатам данного классификатора.
2. Выбрать различное значение порога вероятности отделения классов (не менее 5 различных значений) и вычислить значения метрик качества классификации при различных пороговых значениях. В качестве метрик можно рассматривать: *accuracy*, *precision*, *recall*, *F1-score*.
3. (бонус) по ROC-кривой определить оптимальное значение порога классификации на основе анализа *F1-score*.

Выполненный п.3 добавляет “бонусы в карму”, которые оказывают наибольшее влияние на оценку, но в то же время отсутствие этого пункта не оказывает отрицательного влияния.

## **References:**

🔗 [KNeighborsRegressor: документация](https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html)

🔗 [LogisticRegression: документация](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

🔗 [Пример классификации с различными пороговыми значениями](https://www.geeksforgeeks.org/what-is-the-default-threshold-in-sklearn-logistic-regression/)