**Задание 2.**

Задание по курсу интеллектуальная обработка данных:

1. Загрузить данные в среду Python. Например с помощью

from sklearn.datasets import load\_breast\_cancer

cancer = load\_breast\_cancer()

Данные можно брать и на репозитории

<https://archive.ics.uci.edu/ml/index.php>

**Должна решаться задача классификации**

1. Описать задачу. Что является входом, что выходом, какая задача решается в принципе.
2. Построить классификационную модель с помощью инструмента «к-ближайших соседей»
3. Выбрать наилучшее количество соседей и наилучшую метрику – наилучшее в смысле наименьшей ошибке на тестовом множестве. Использовать специальную процедуру поиска. Привести таблицу запусков:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Количество соседей =1 | Количество соседей =2 | Количество соседей =3 | … |
| Метрика 1 | Среднее значение по результатам кросс-валидации |  |  |  |
| Метрика 2 |  |  |  |  |
| …. |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Выбрать параметры с наибольшим значением средней точности

1. Для выбранных параметров получить выход модели и сравнить с выходом реальным с помощью графического представления результата
2. Сделать выводы о применимости данной модели к этой задаче.
3. Оформить отчет.

**Задание 3.**

Выполнить то же самое, но для задачи **регрессии**