МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Лабораторная работа №4**

по курсу Методы машинного обучения

«Подготовка обучающей и тестовой выборки, кросс-валидация и подбор гиперпараметров на примере метода ближайших соседей»

17

(количество листов)

|  |  |
| --- | --- |
| ИСПОЛНИТЕЛЬ: |  |
| студент группы ИУ5-24М | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Савельева М. А. | "\_\_"\_\_\_\_\_\_\_\_\_2019 г. |

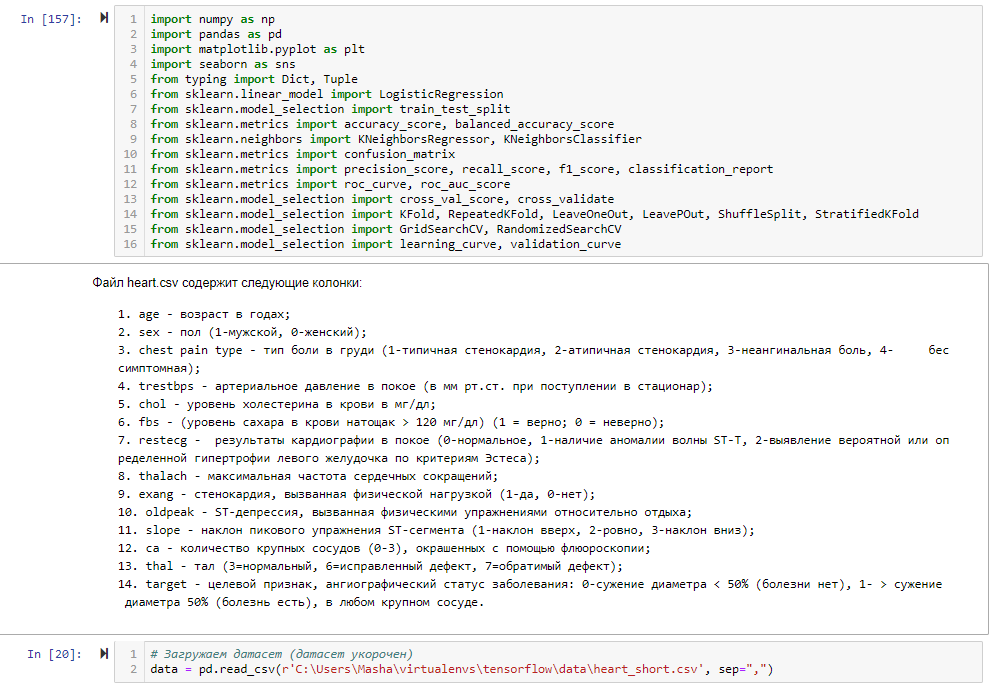
|  |  |
| --- | --- |
| ПРЕПОДАВАТЕЛЬ: |  |
| Гапанюк Ю. Е.  к.т.н., доцент | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  | "\_\_"\_\_\_\_\_\_\_\_\_2019 г. |

Москва – 2019

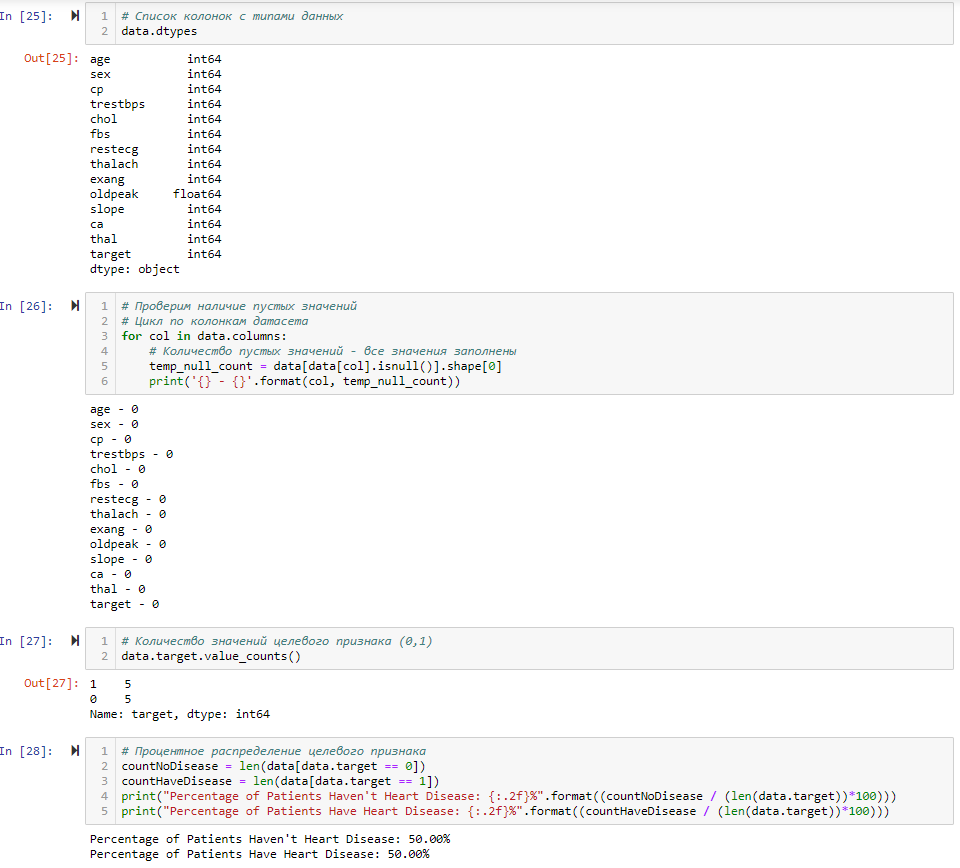
**Задание**

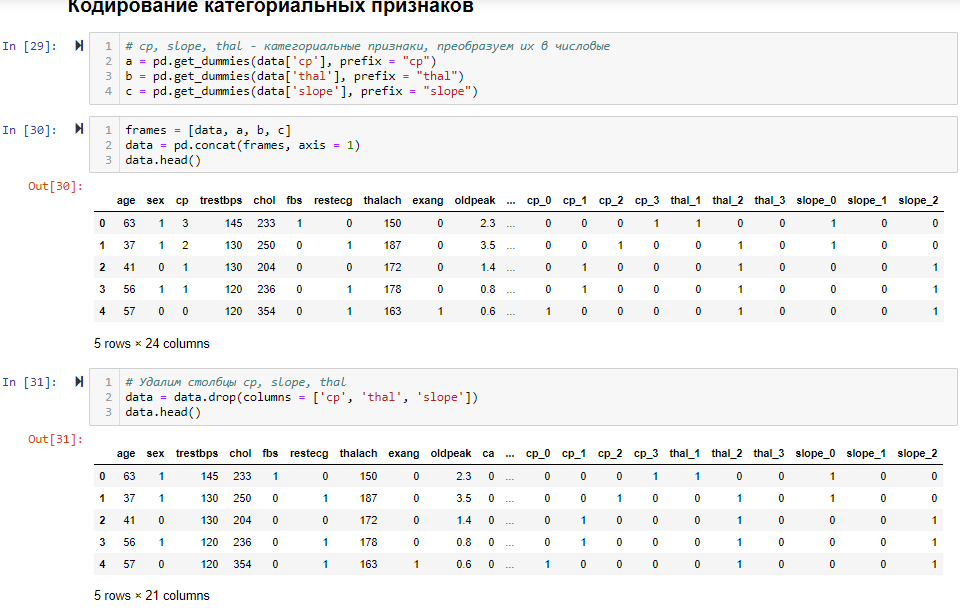
1. Выберите набор данных (датасет) для решения задачи классификации или регресии.
2. В случае необходимости проведите удаление или заполнение пропусков и кодирование категориальных признаков.
3. С использованием метода train\_test\_split разделите выборку на обучающую и тестовую.
4. Обучите модель ближайших соседей для произвольно заданного гиперпараметра K. Оцените качество модели с помощью трех подходящих для задачи метрик.
5. Постройте модель и оцените качество модели с использованием кросс-валидации. Проведите эксперименты с тремя различными стратегиями кросс-валидации.
6. Произведите подбор гиперпараметра K с использованием GridSearchCV и кросс-валидации.
7. Повторите пункт 4 для найденного оптимального значения гиперпараметра K. Сравните качество полученной модели с качеством модели, полученной в пункте 4.
8. Постройте кривые обучения и валидации.

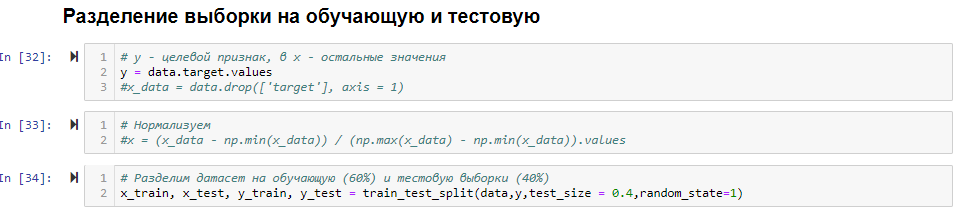
**Текст программы**

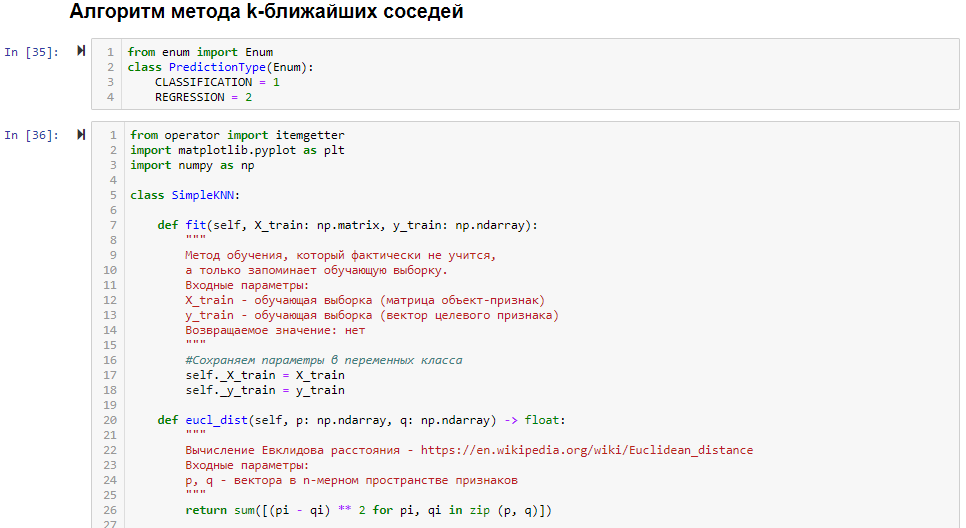






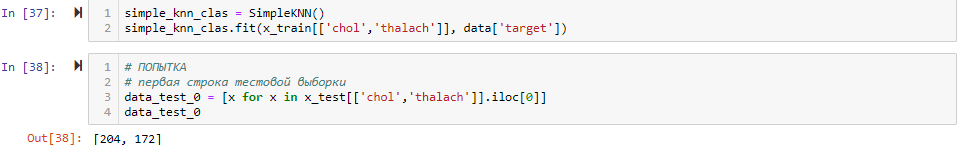


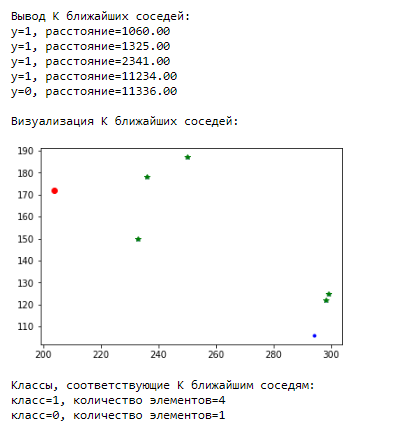
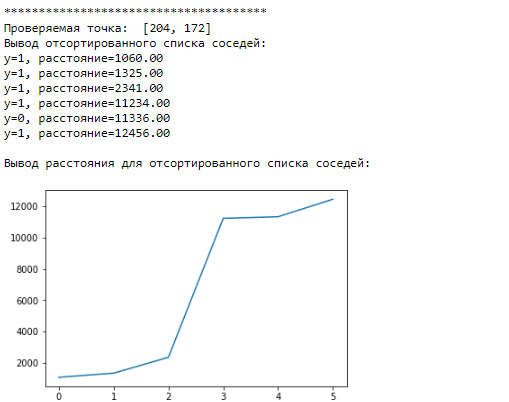
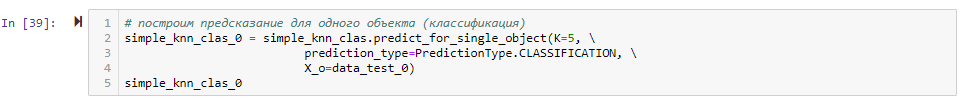


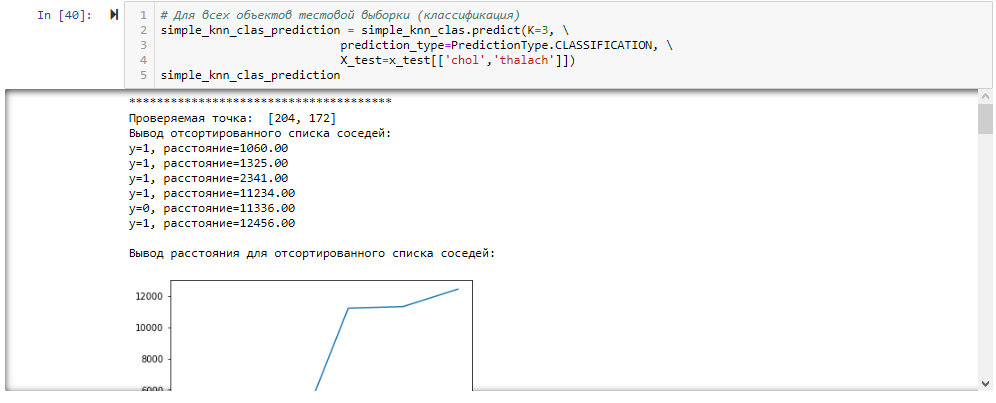


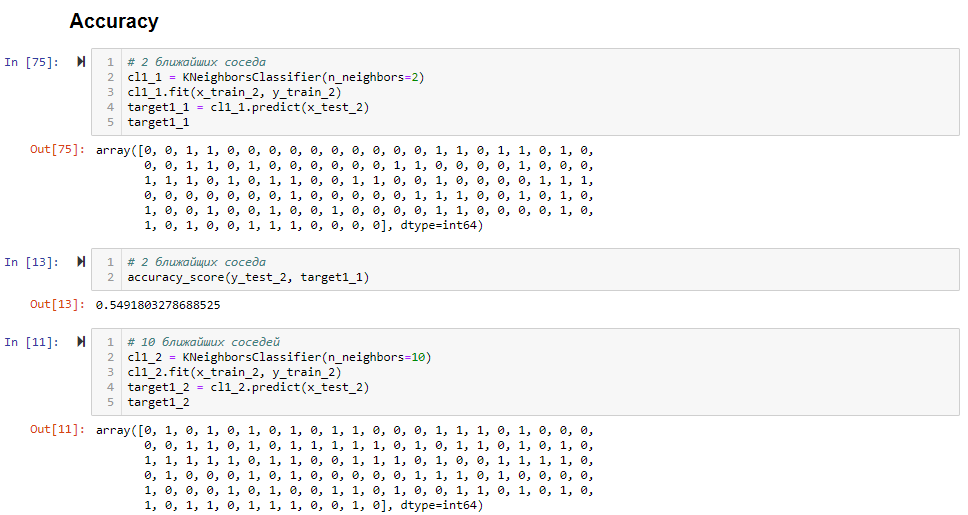
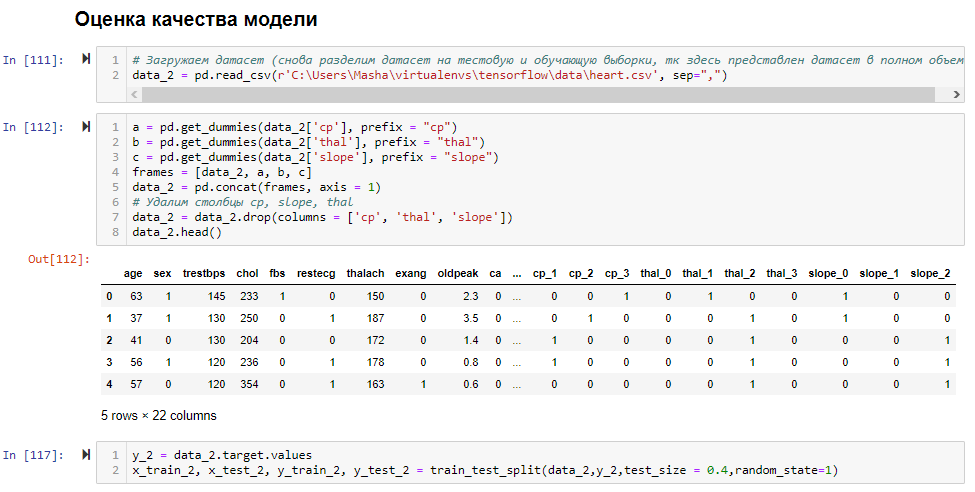


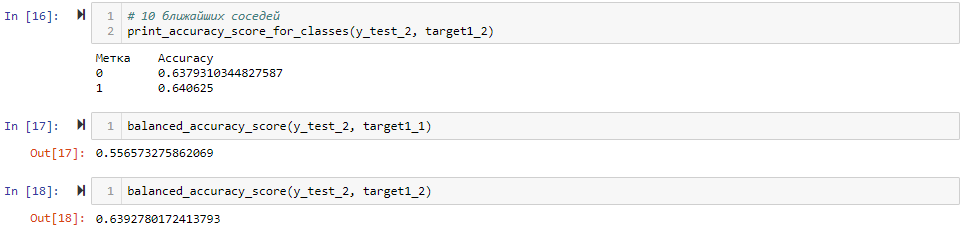
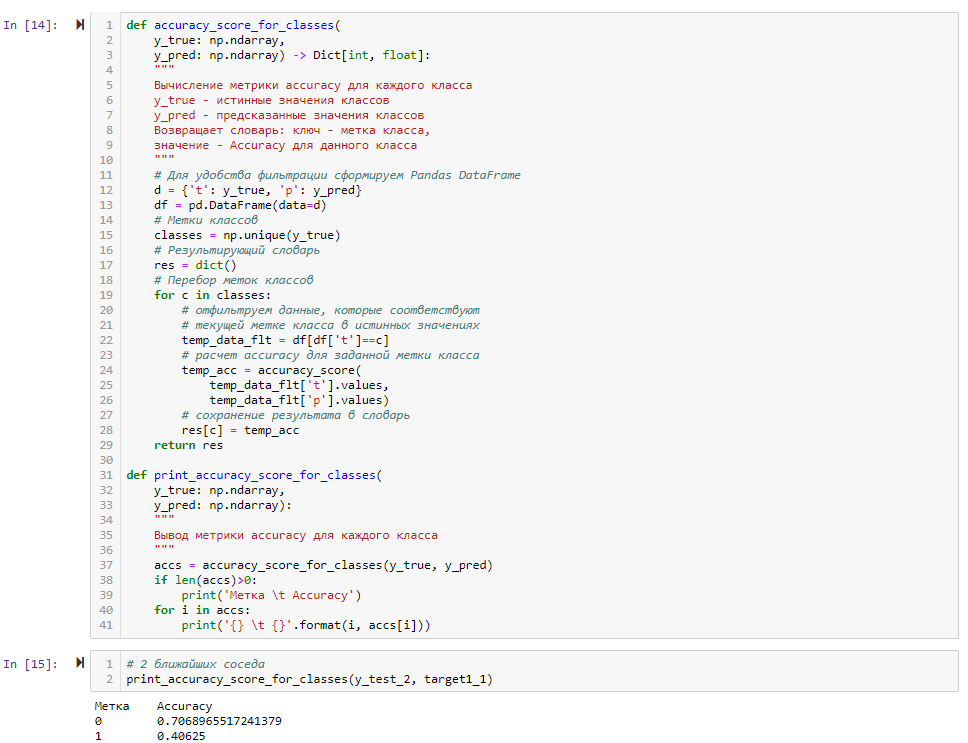


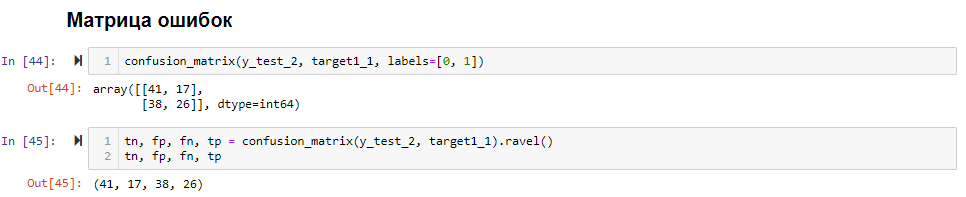


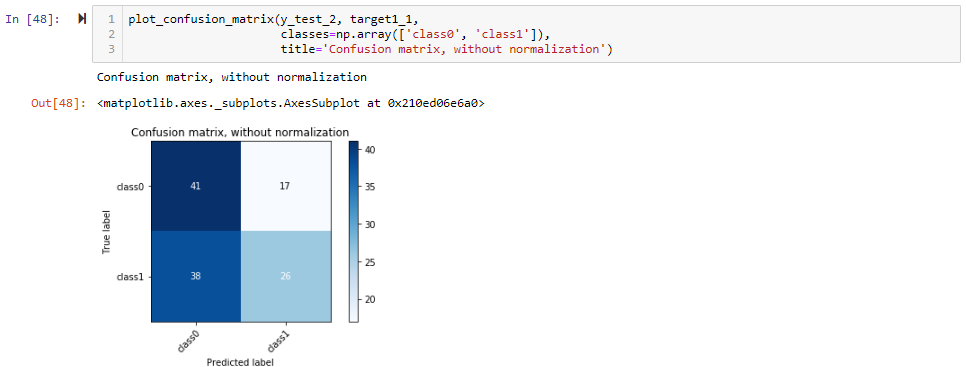


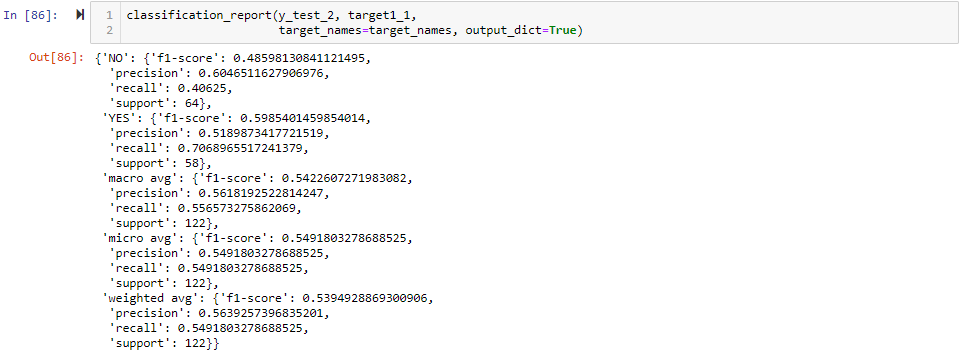
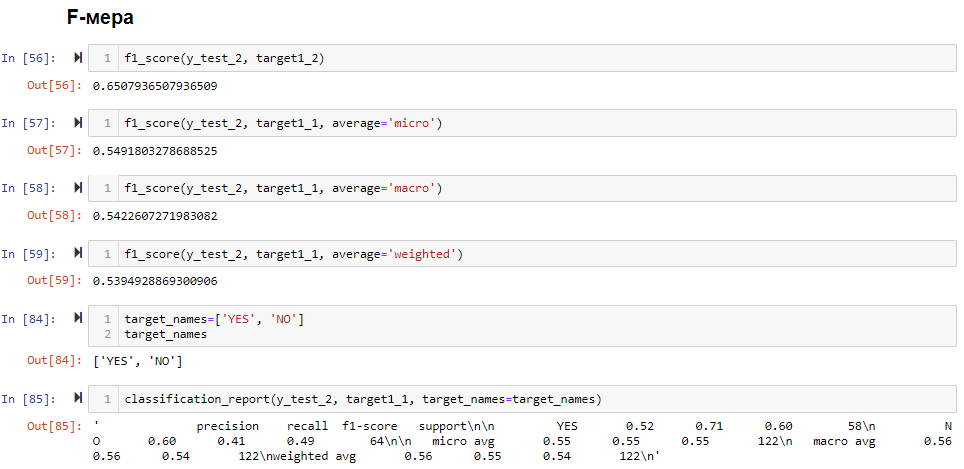


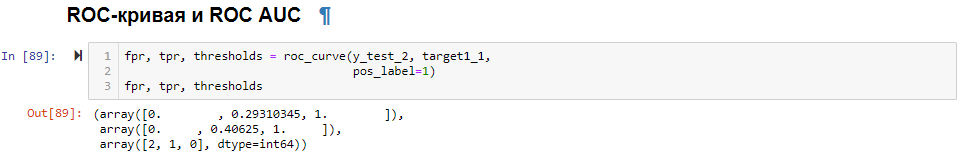


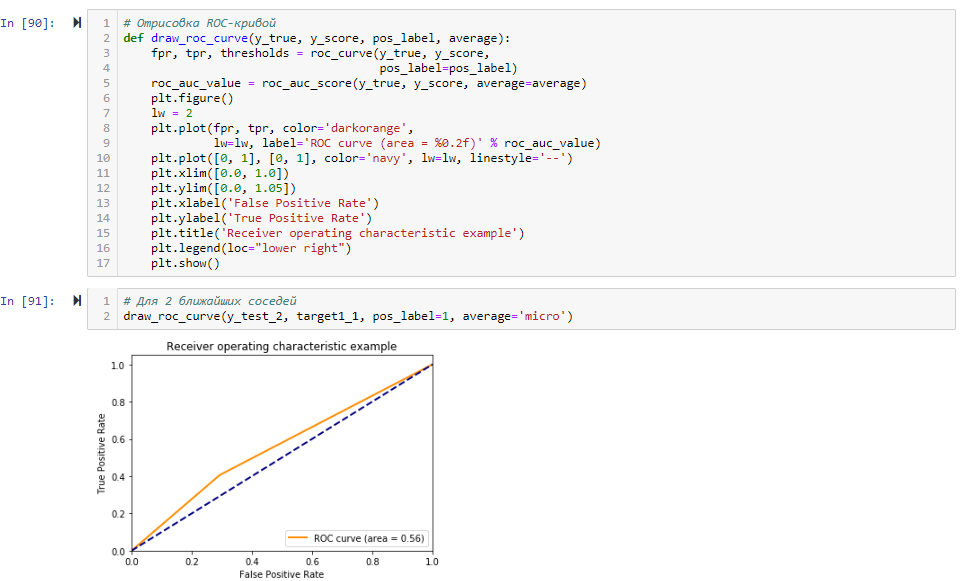


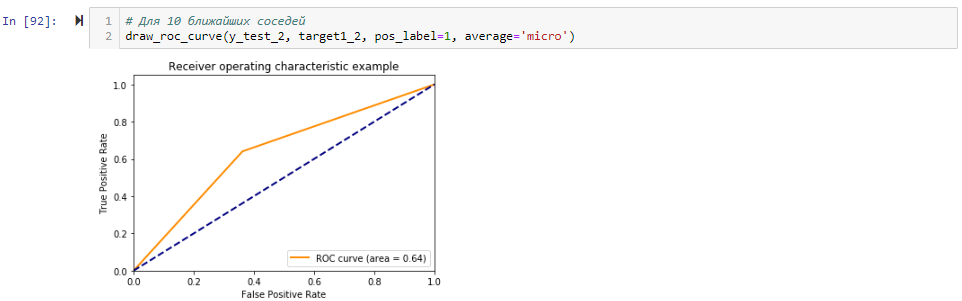




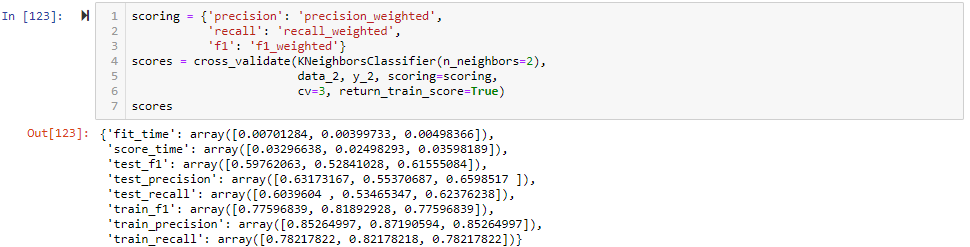




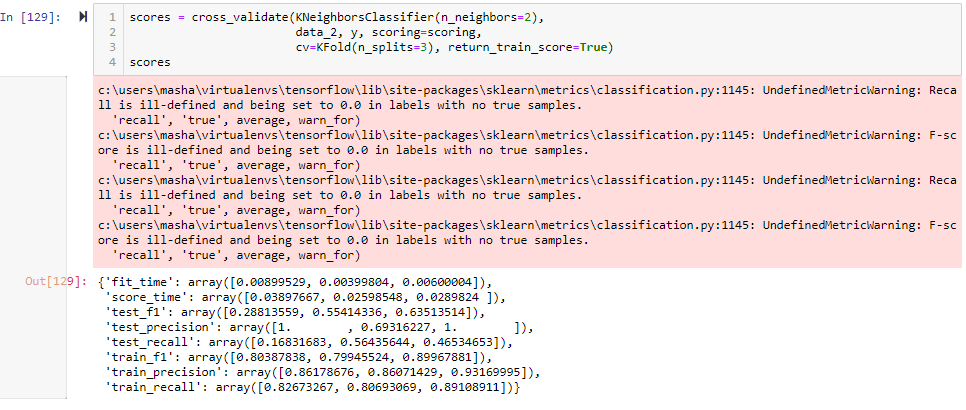


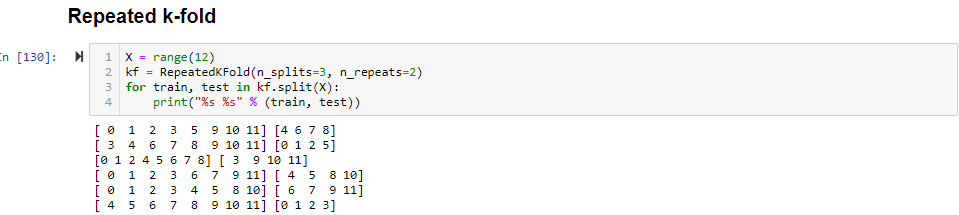


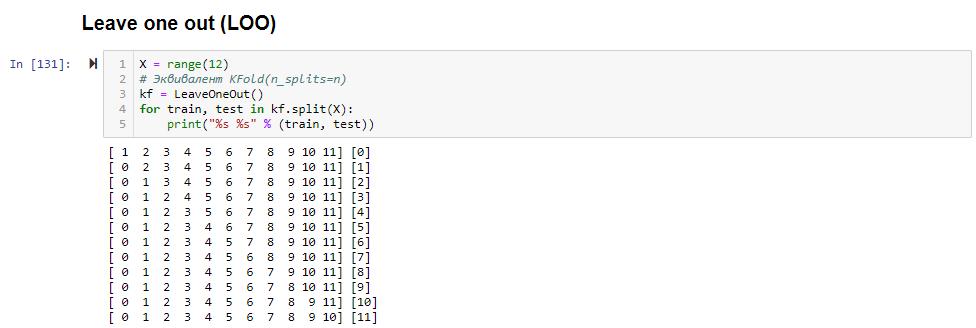


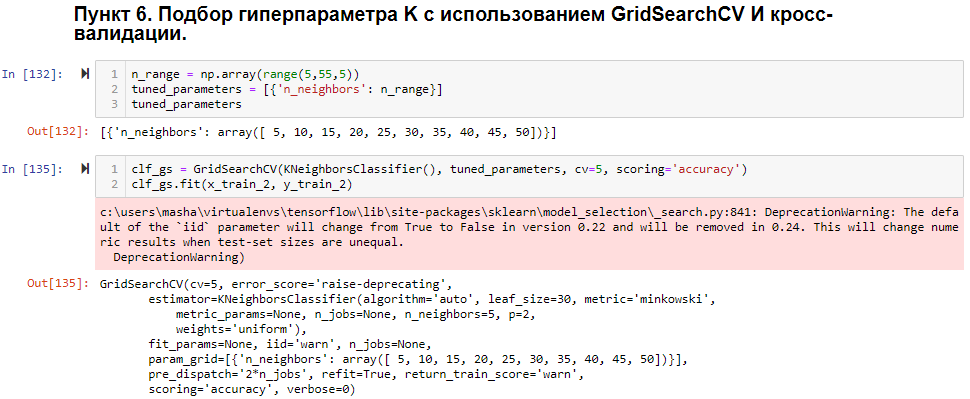


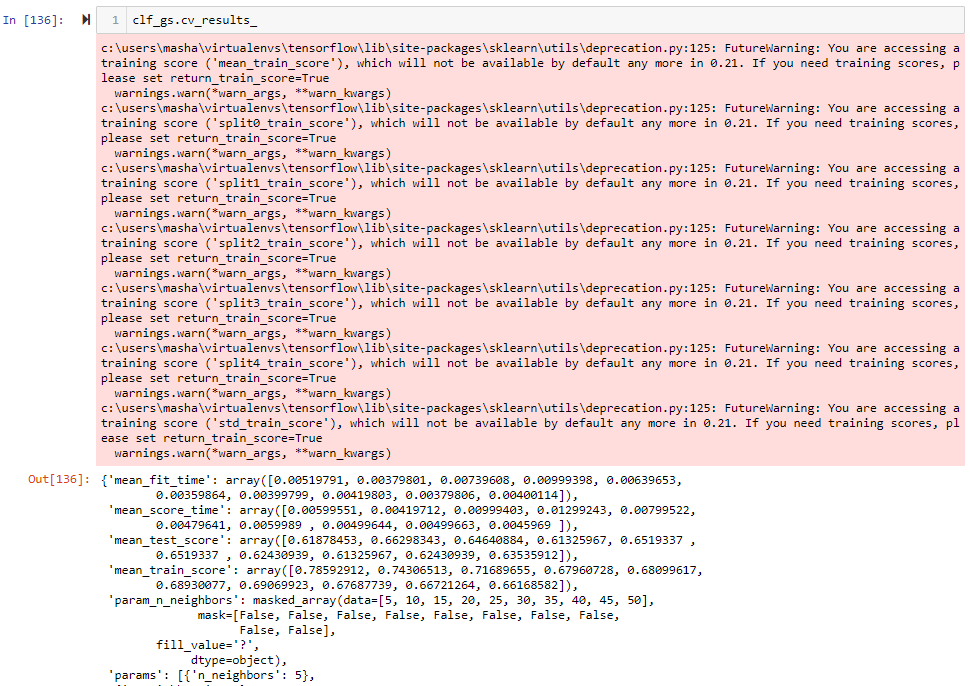


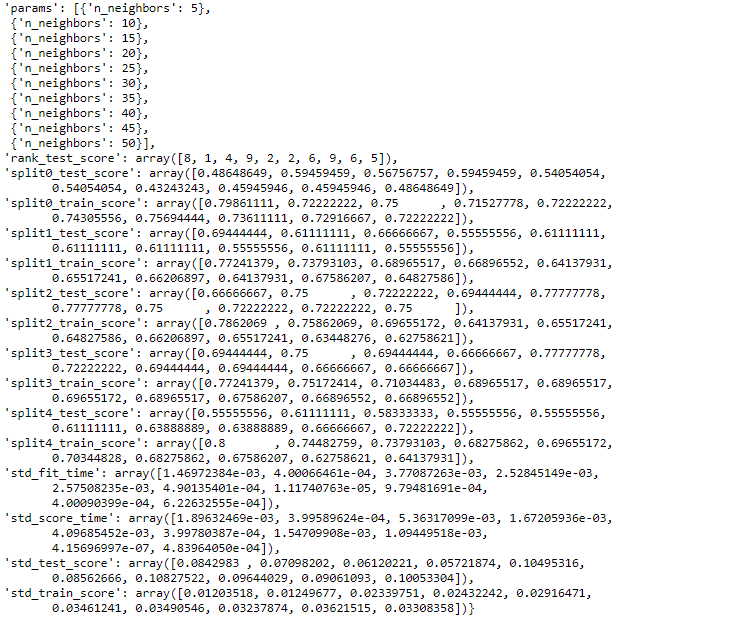


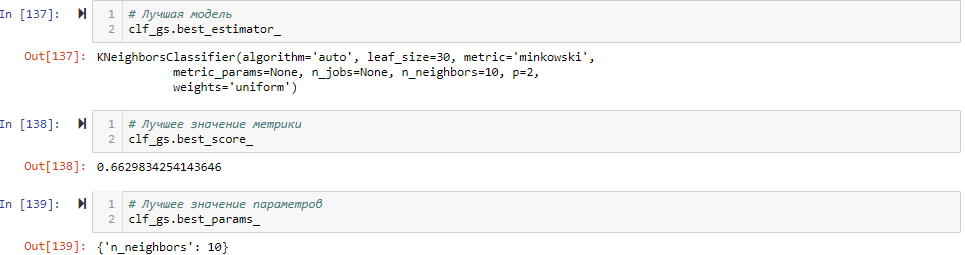


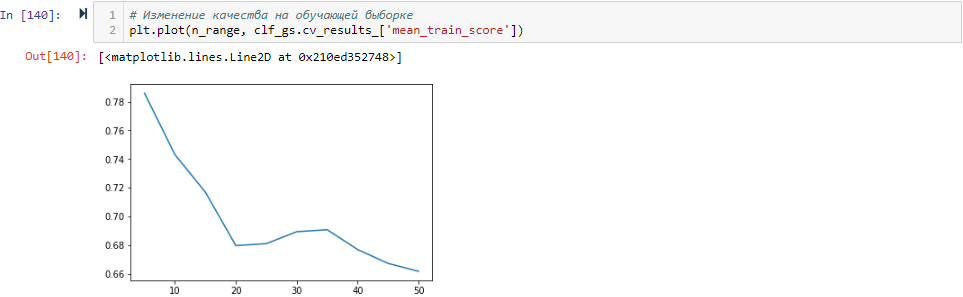


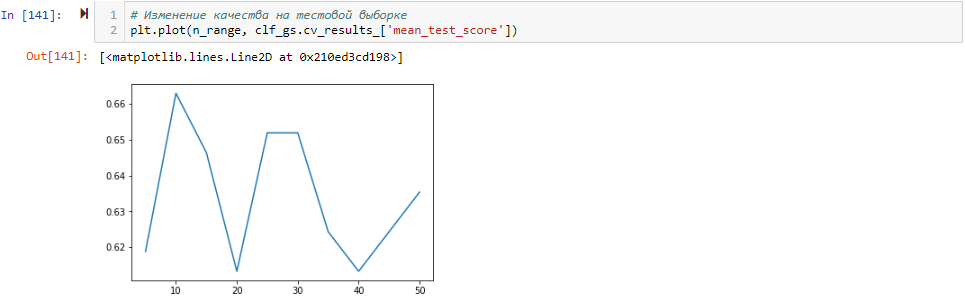


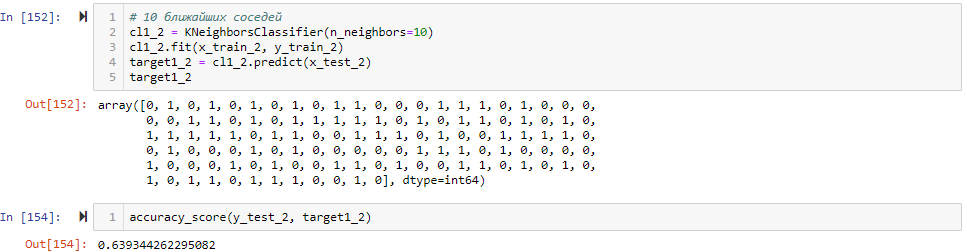




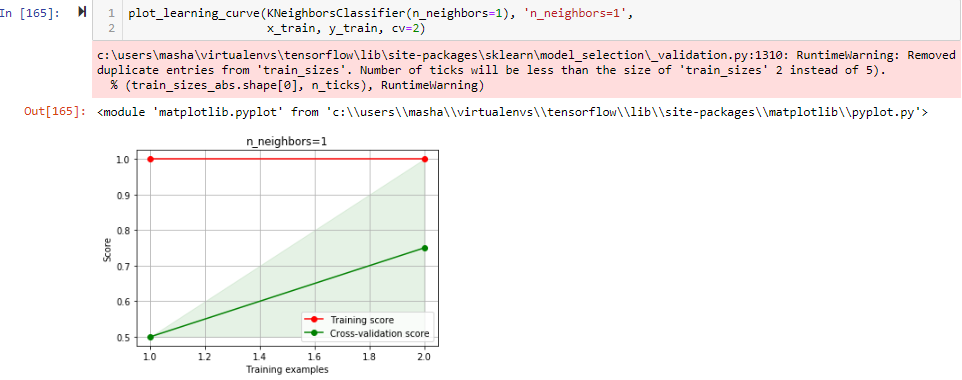


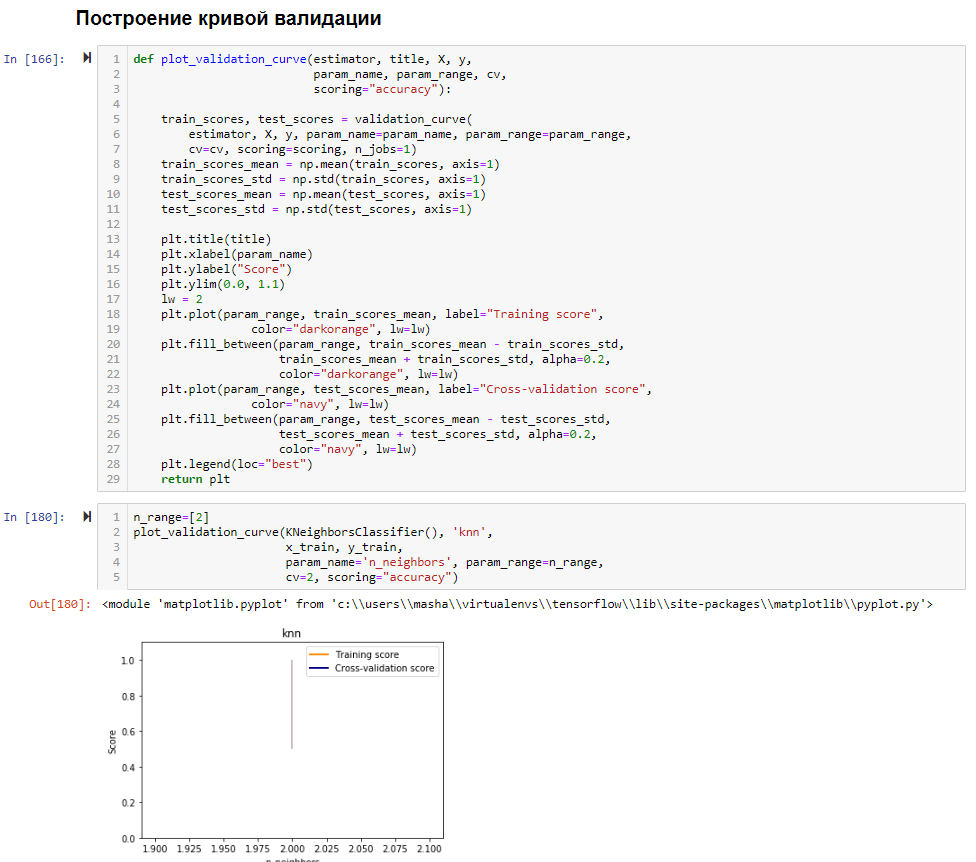












**Вывод**

В результате проведения лабораторной работы были получены навыки разделения выборки на обучающую и тестовую. Была построена модель обучения по методу ближайших соседей для произвольного заданного K параметра, было оценено качество построенной модели, в том числе с помощью кросс-валидации. Был произведен подбор параметра K с использованием GridSearchCV. Построены кривые обучения и валидации.