|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

**Факультет «Информатика и системы управления»**

**Кафедра ИУ5 «Системы обработки информации и управления»**

Курсовая работа

по дисциплине «Технологии машинного обучения»

на тему:

Исследование датасета. Решение задачи классификации.

Выполнил:

студент группы № ИУ5-62

Король К.Э.

подпись, дата

Проверил:

Гапанюк Ю.Е.

подпись, дата

2019 г.

**Задание**

1. Поиск и выбор набора данных для построения моделей машинного обучения. На основе выбранного набора данных студент должен построить модели машинного обучения для решения или задачи классификации, или задачи регрессии.
2. Проведение разведочного анализа данных. Построение графиков, необходимых для понимания структуры данных. Анализ и заполнение пропусков в данных.
3. Выбор признаков, подходящих для построения моделей. Кодирование категориальных признаков Масштабирование данных. Формирование вспомогательных признаков, улучшающих качество моделей.
4. Проведение корреляционного анализа данных. Формирование промежуточных выводов о возможности построения моделей машинного обучения. В зависимости от набора данных, порядок выполнения пунктов 2, 3, 4 может быть изменен.
5. Выбор метрик для последующей оценки качества моделей. Необходимо выбрать не менее трех метрик и обосновать выбор.
6. Выбор наиболее подходящих моделей для решения задачи классификации или регрессии. Необходимо использовать не менее пяти моделей, две из которых должны быть ансамблевыми.
7. Формирование обучающей и тестовой выборок на основе исходного набора данных.
8. Построение базового решения (baseline) для выбранных моделей без подбора гиперпараметров. Производится обучение моделей на основе обучающей выборки и оценка качества моделей на основе тестовой выборки.
9. Подбор гиперпараметров для выбранных моделей. Рекомендуется использовать методы кросс-валидации. В зависимости от используемой библиотеки можно применять функцию GridSearchCV, использовать перебор параметров в цикле, или использовать другие методы.
10. Повторение пункта 8 для найденных оптимальных значений гиперпараметров. Сравнение качества полученных моделей с качеством baseline-моделей.
11. Формирование выводов о качестве построенных моделей на основе выбранных метрик. Результаты сравнения качества рекомендуется отобразить в виде графиков и сделать выводы в форме текстового описания. Рекомендуется постройение графиков обучения и валидации, влияния значений гиперпарметров на качество моделей и т.д.

**Оглавление**

[Введение 4](#_Toc10193111)

[Основная часть 5](#_Toc10193112)

[Заключение 2](#_Toc10193113)3

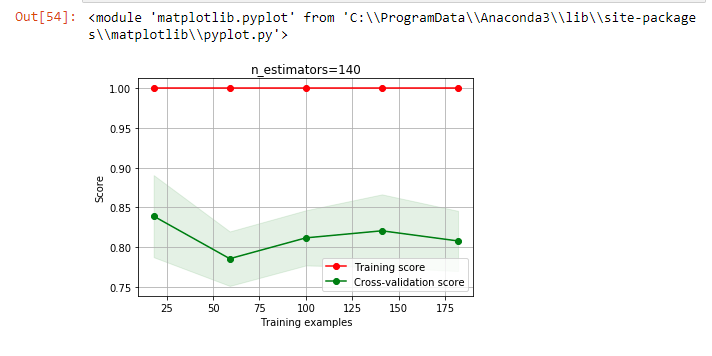
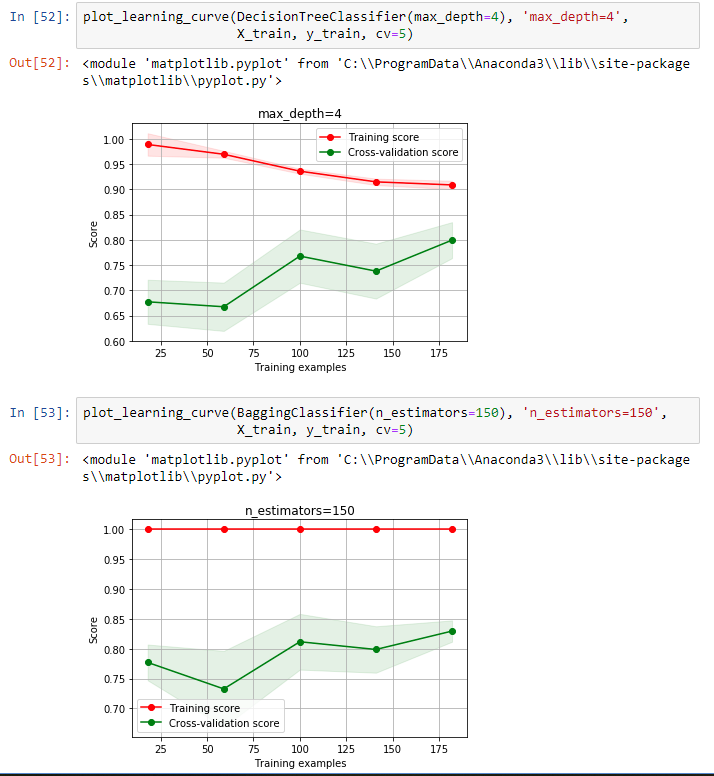
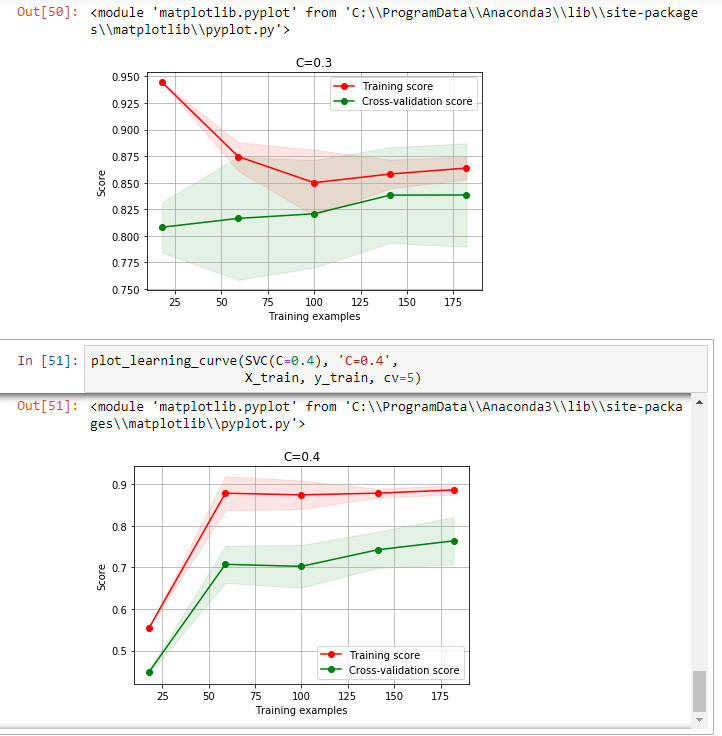
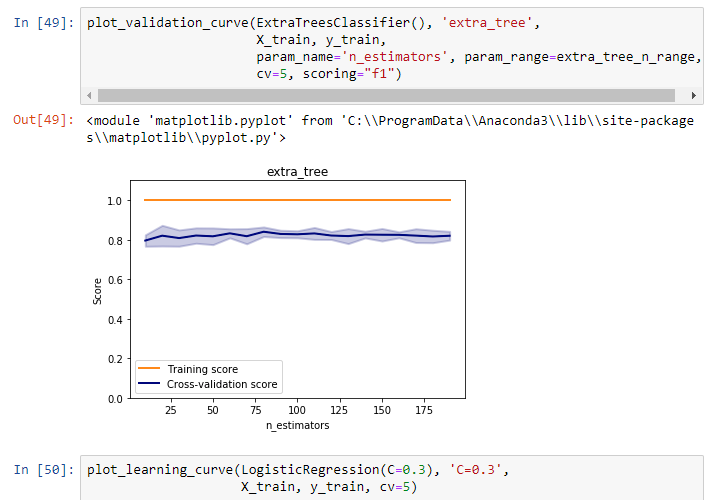
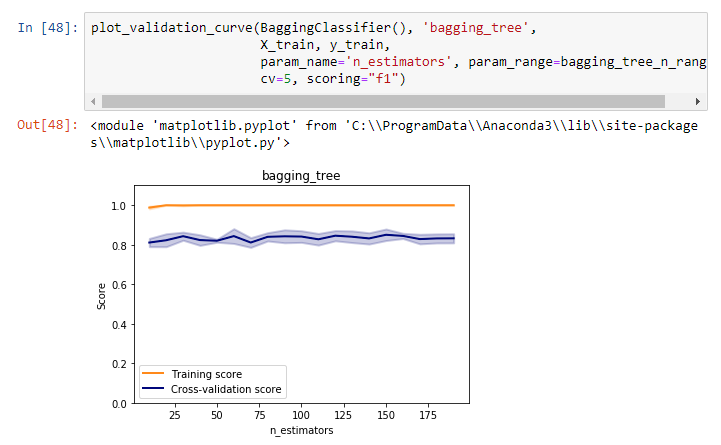
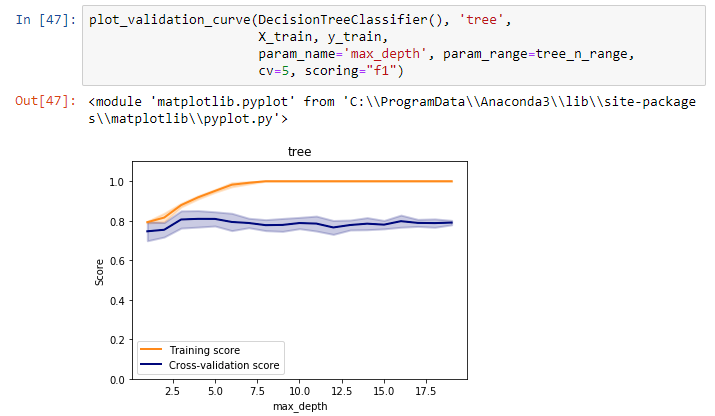
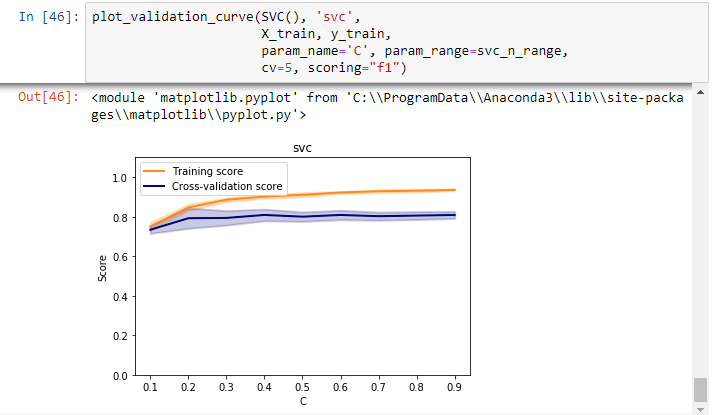
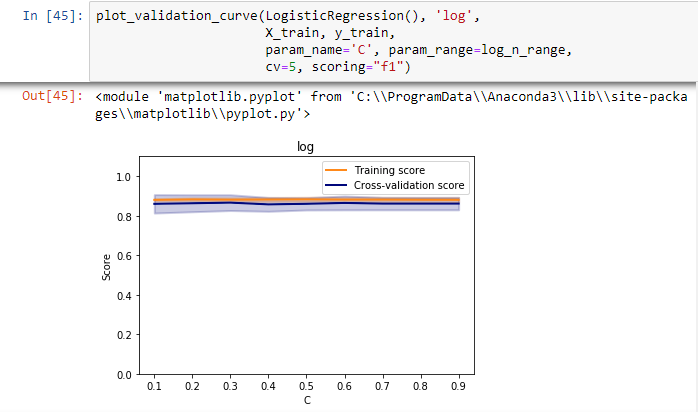
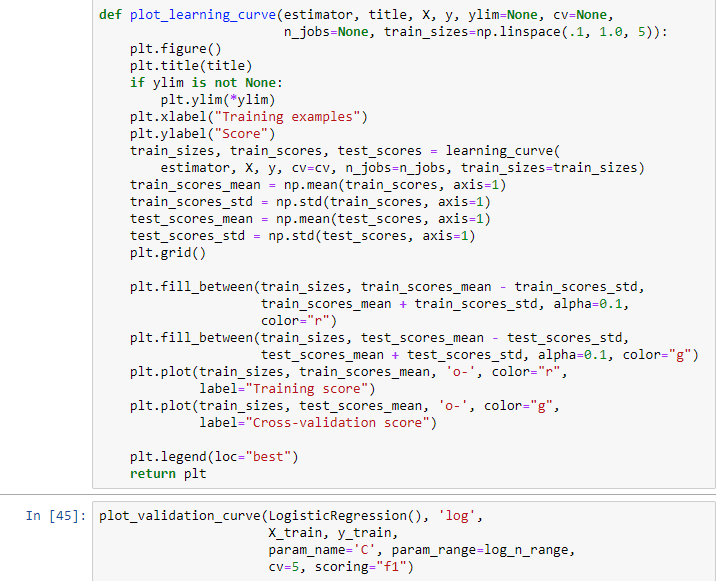
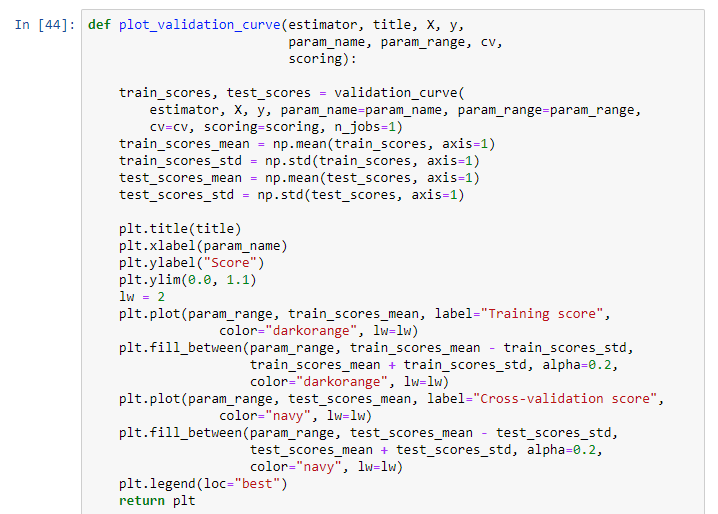
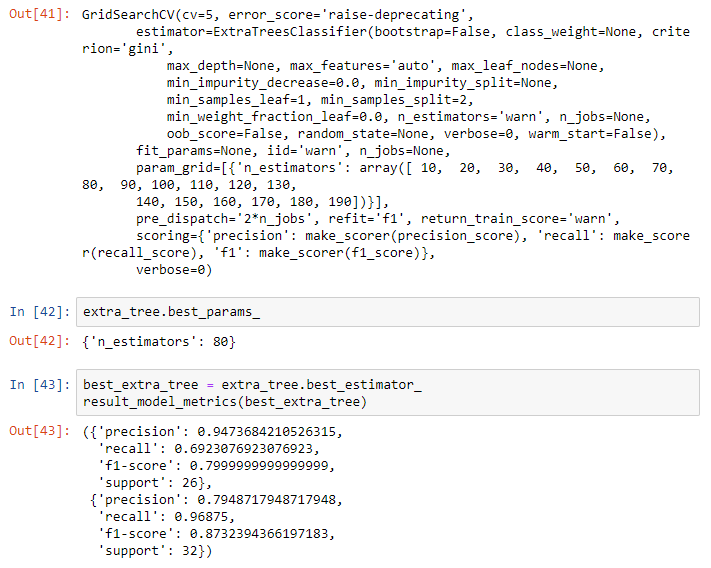
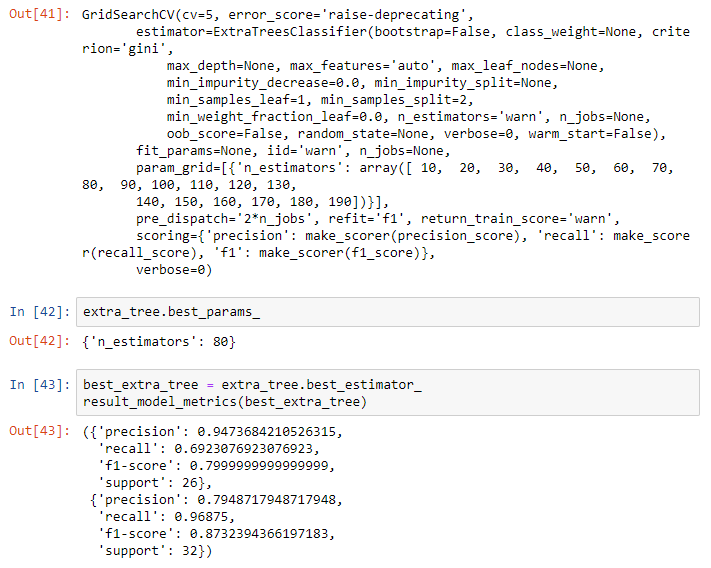
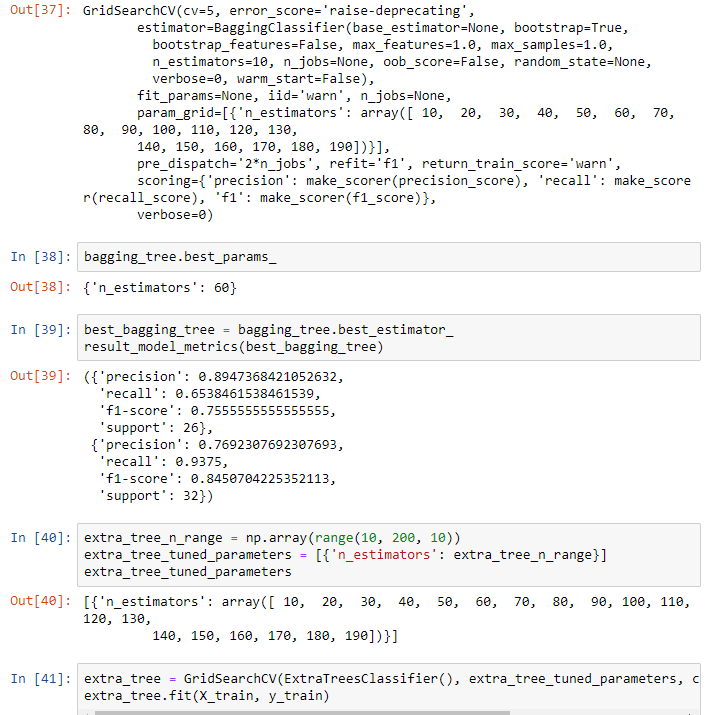
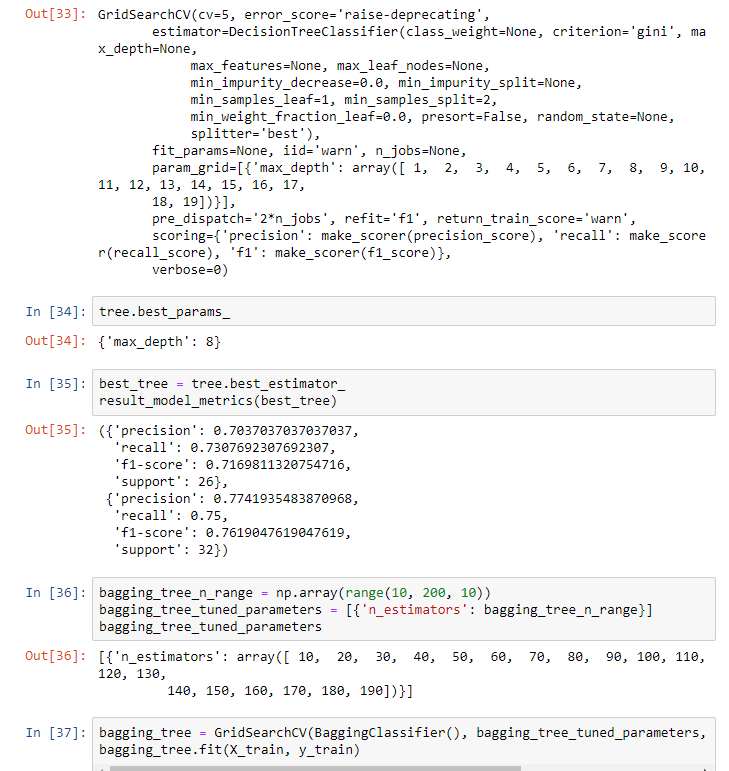
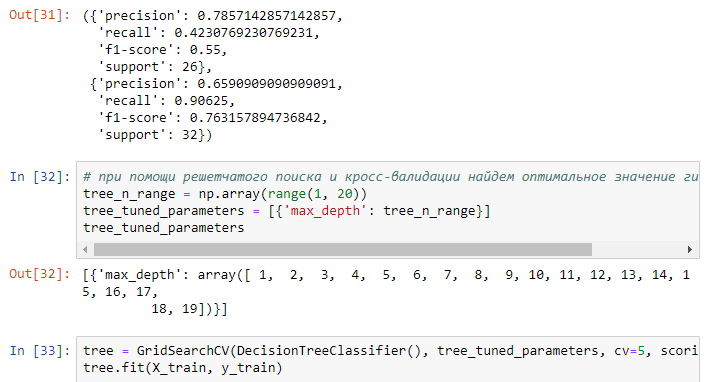
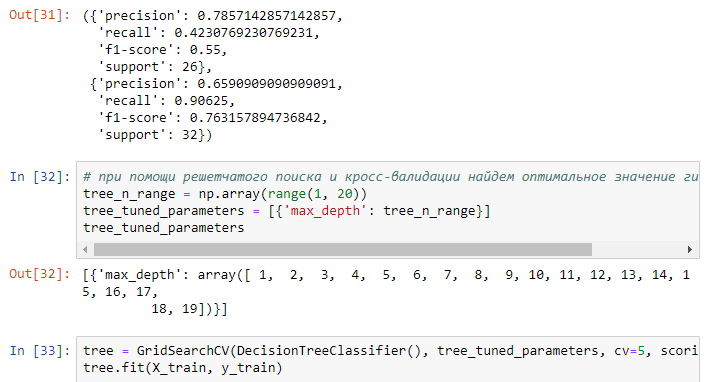
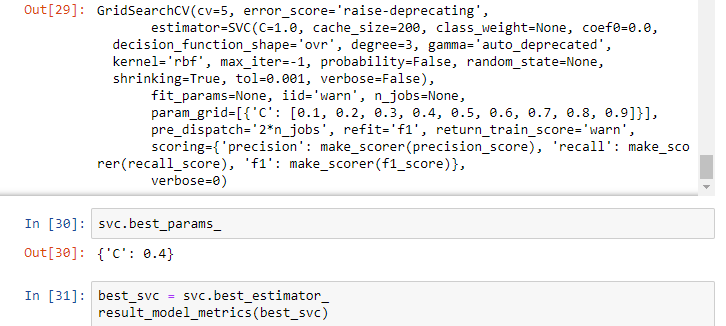
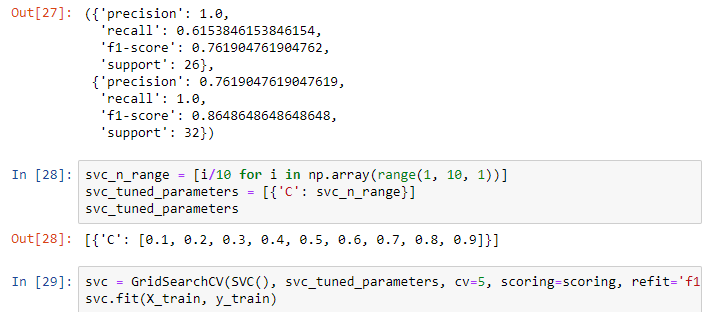
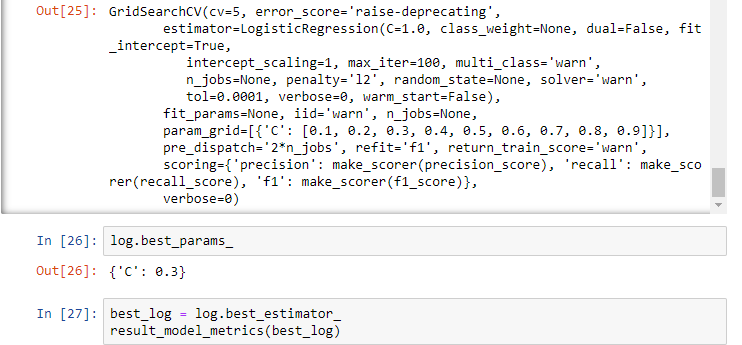
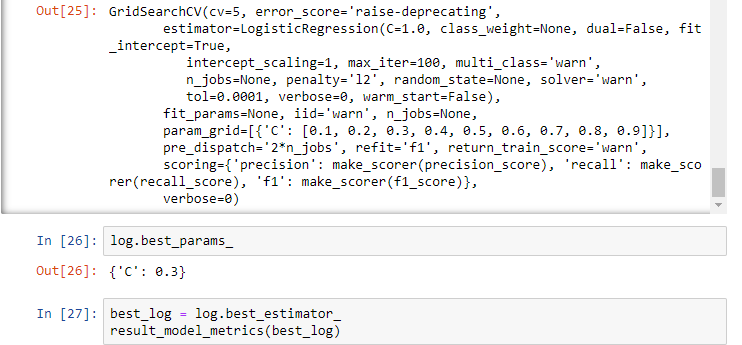
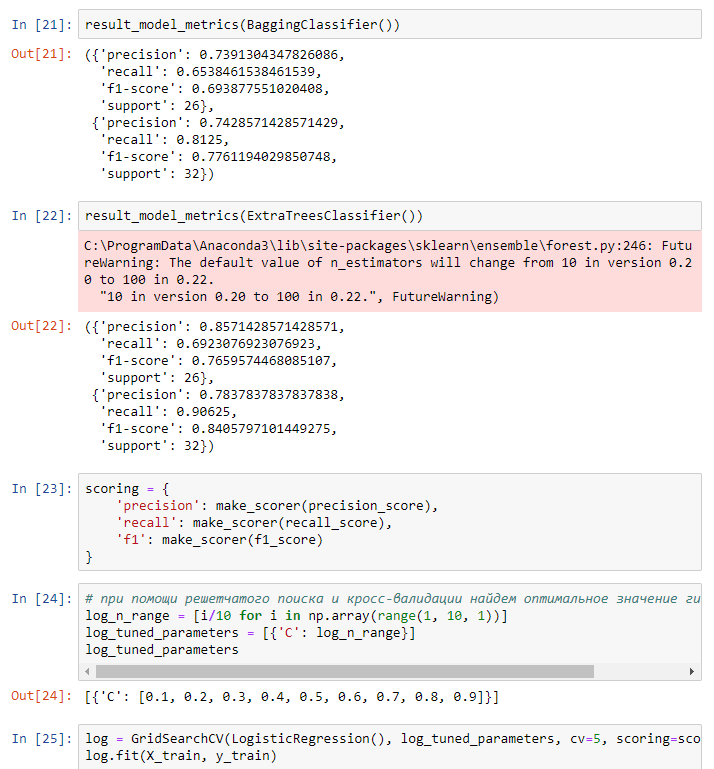
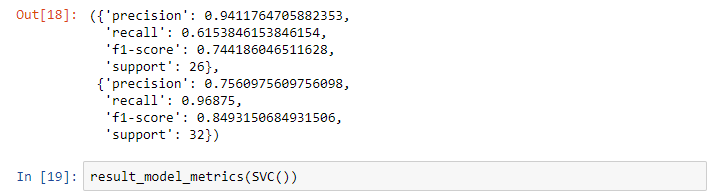
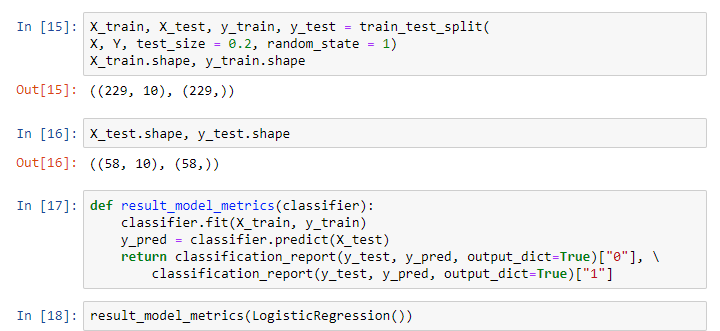
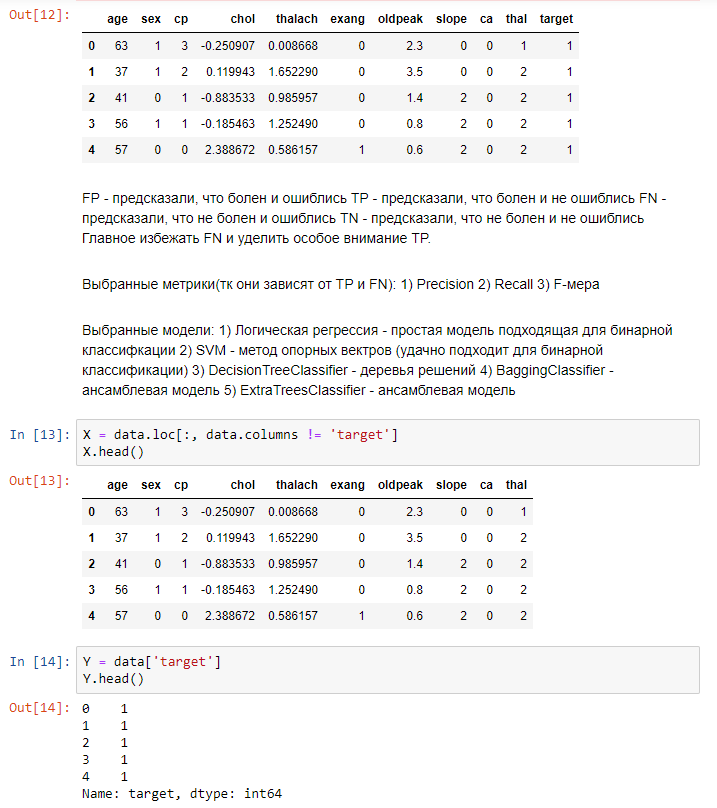
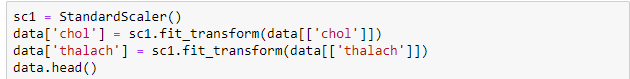
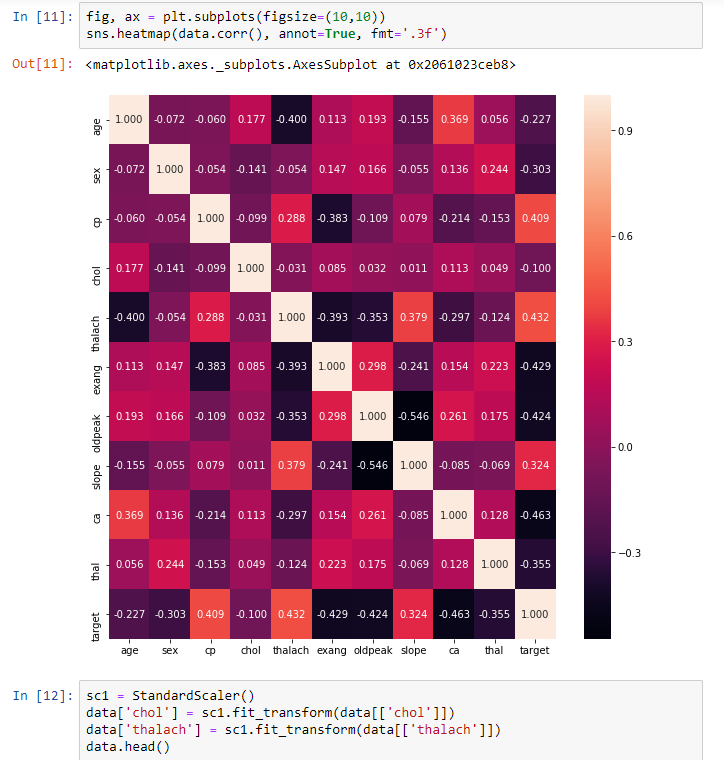
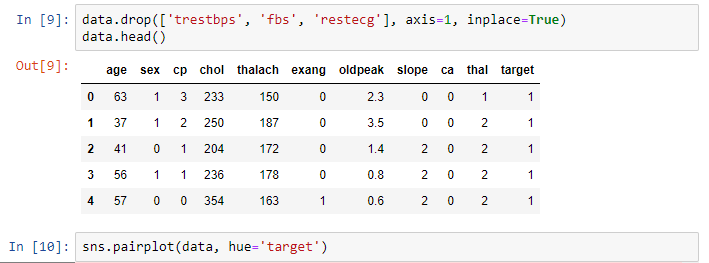
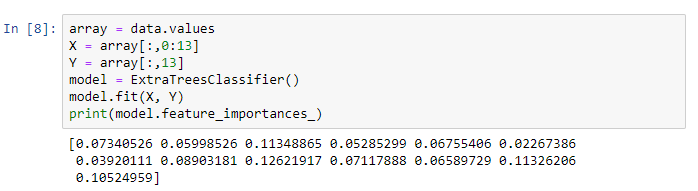
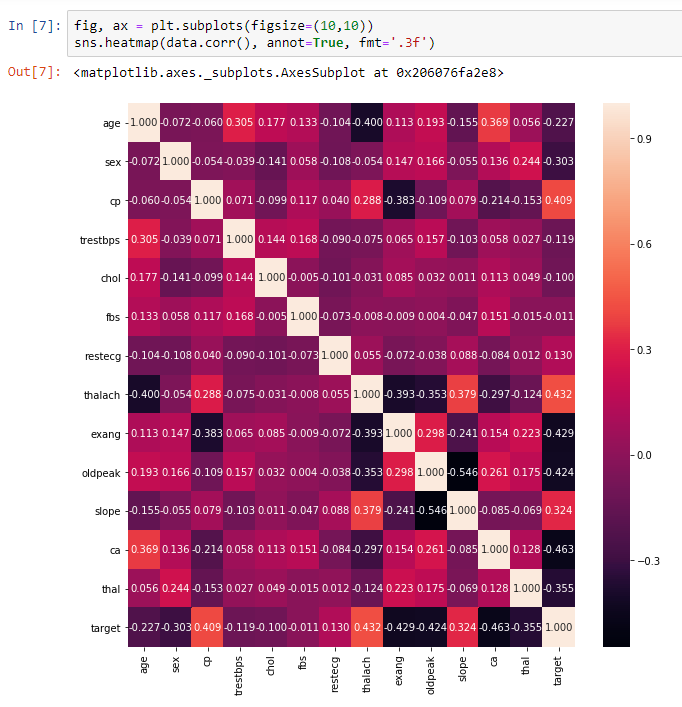
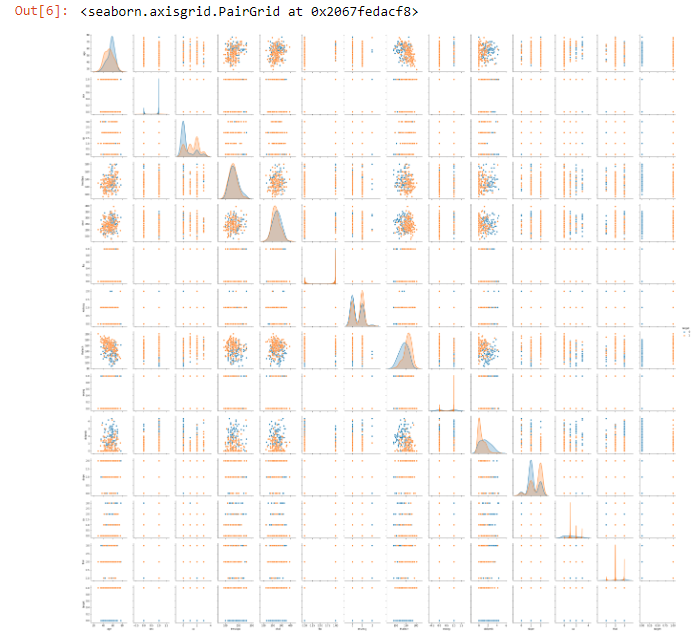
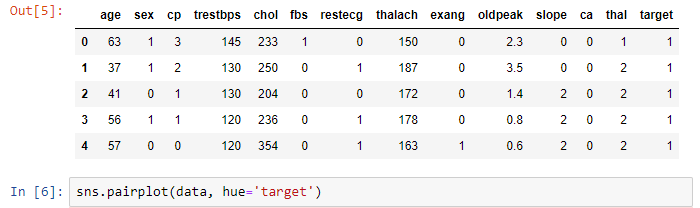
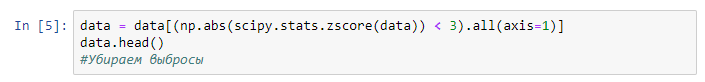
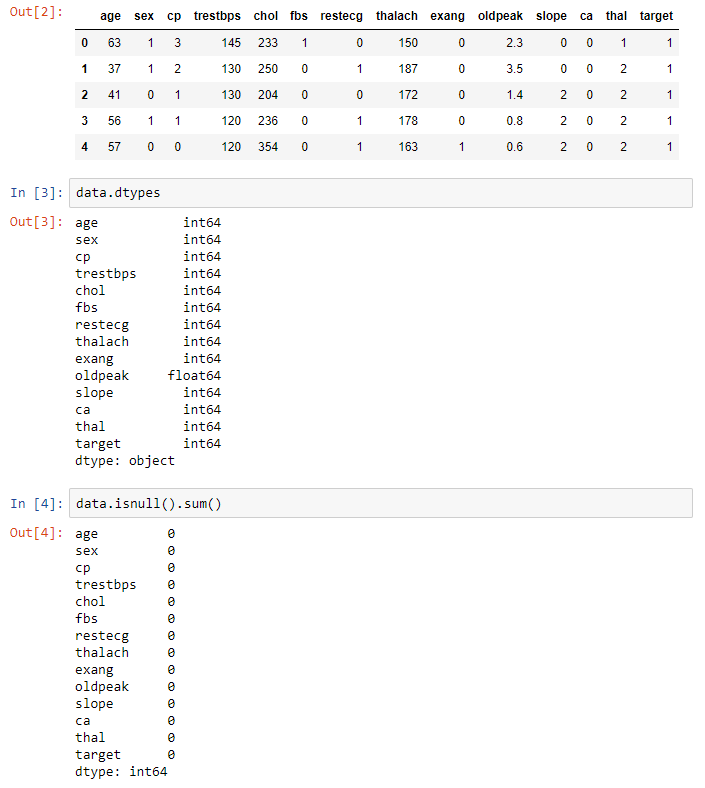
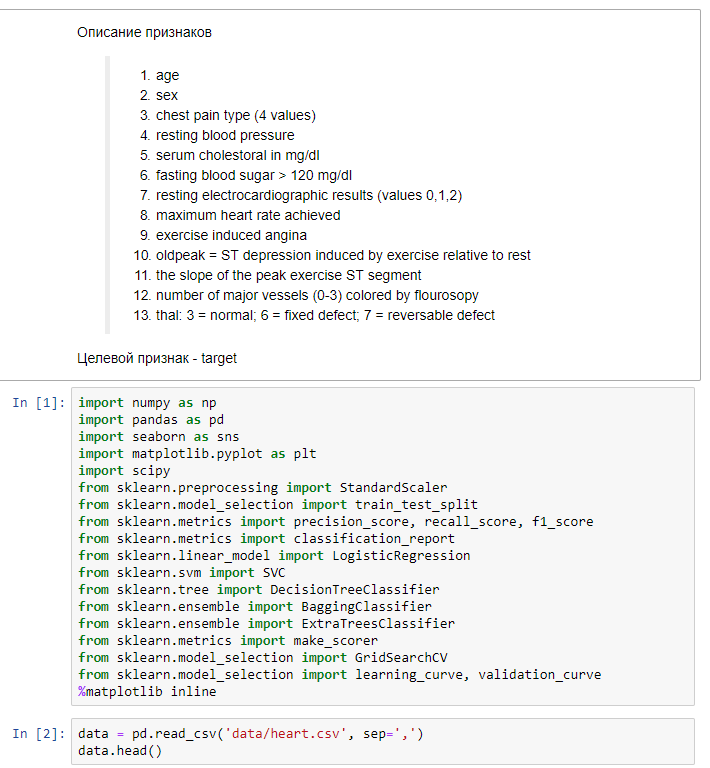
[Список использованной литературы 24](#_Toc10193114)

# 

# Введение

Данная работа предназначена для усвоения и закрепления знаний по дисциплине «Технологии машинного обучения». Здесь закрепляются навыки проведения разведочного анализа данных, выбора признаков для построения модели, проведения корреляционного анализа, подбора метрик, решения задачи классификации, построения базового решения и подбора гиперпараметров.

# Основная часть



В итоге получилось улучшить результаты классификации на основе выбранных метрик после подбора гиперпарметров для используемых моделей.

# Заключение

В данной работе были закреплены навыки проведения разведочного анализа данных, выбора признаков для построения модели, проведения корреляционного анализа, подбора метрик, решения задачи классификации, построения базового решения и подбора гиперпараметров, сравнены качества моделей.

# Список использованной литературы

# <https://github.com/ugapanyuk/ml_course/wiki/COURSE_TMO>

# <https://habr.com/ru/company/ods/blog/328372/>

# <https://habr.com/ru/company/nixsolutions/blog/425253/>

# [<https://habr.com/ru/company/nixsolutions/blog/425907/>](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html)