|  |  |
| --- | --- |
| lu135925on3bu_tmp_3360867a00ce4d37 | **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 и искусственный интеллект

КАФЕДРА                  Системы обработки информации и управления

**Лабораторная работа №3** **по курсу «Методы машинного обучения в автоматизированных системах обработки информации и управления»**

Подготовили:

**У Жун**

**ИУ5И-25М**

20.05.2024

Проверил:

**Гапанюк Ю. Е.**

*2024 г*.

**Задание:**

1. Выбрать один или несколько наборов данных (датасетов) для решения следующих задач. Каждая задача может быть решена на отдельном датасете, или несколько задач могут быть решены на одном датасете. Просьба не использовать датасет, на котором данная задача решалась в лекции.
2. Для выбранного датасета (датасетов) на основе материалов лекций решить следующие задачи:
   1. масштабирование признаков (не менее чем тремя способами);
   2. обработку выбросов для числовых признаков (по одному способу для удаления выбросов и для замены выбросов);
   3. обработку по крайней мере одного нестандартного признака (который не является числовым или категориальным);
   4. отбор признаков:
      * один метод из группы методов фильтрации (filter methods);
      * один метод из группы методов обертывания (wrapper methods);
      * один метод из группы методов вложений (embedded methods).
3. Загрузка наборов данных и информации



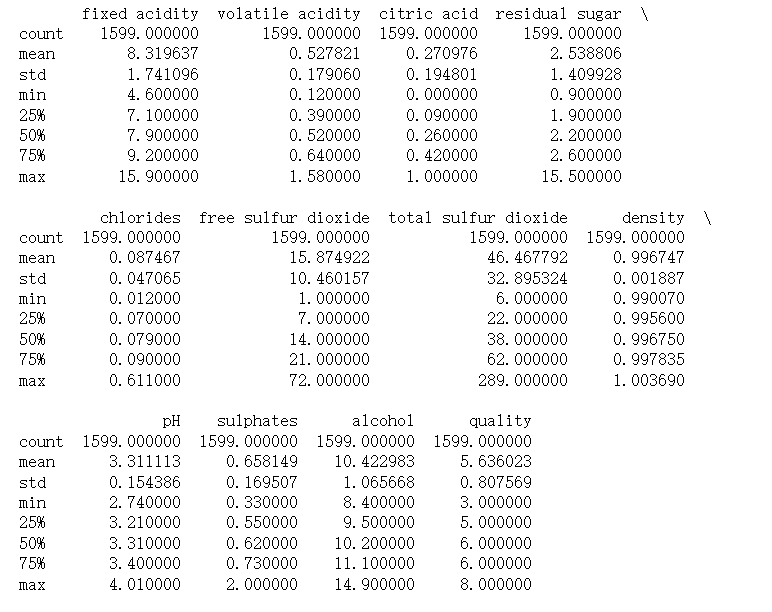
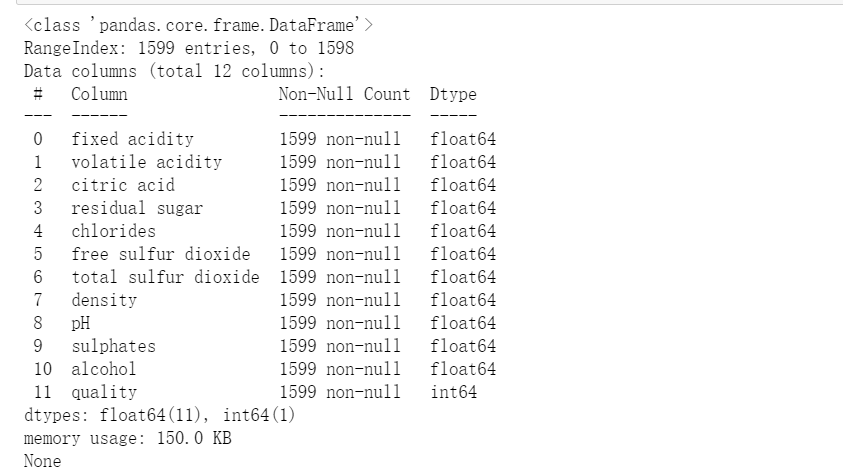


Рис.1-Загрузка наборов данных и информации

1. Масштабирование признаков

Я использую три метода масштабирования: StandardScaler, MinMaxScaler и RobustScaler.

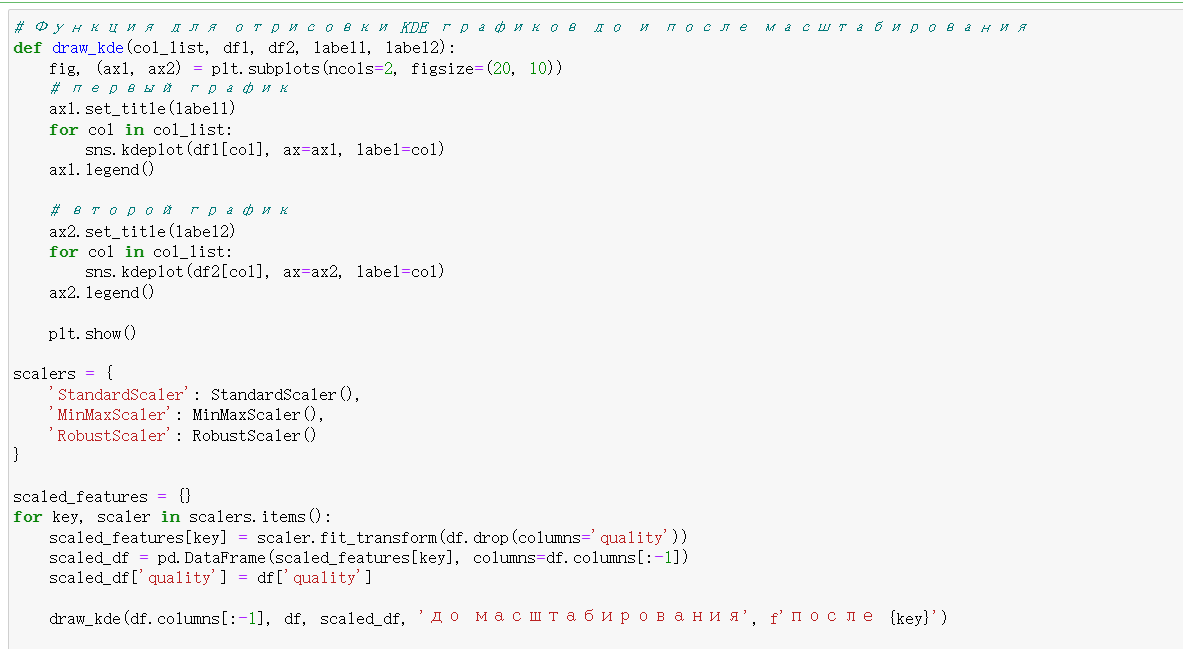
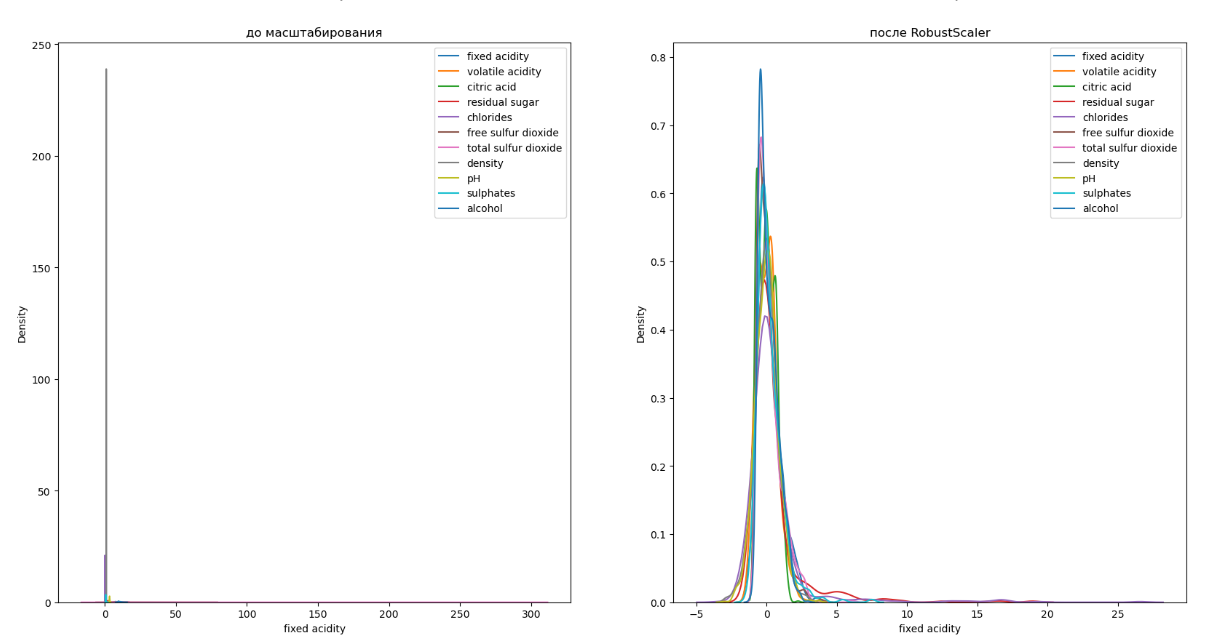
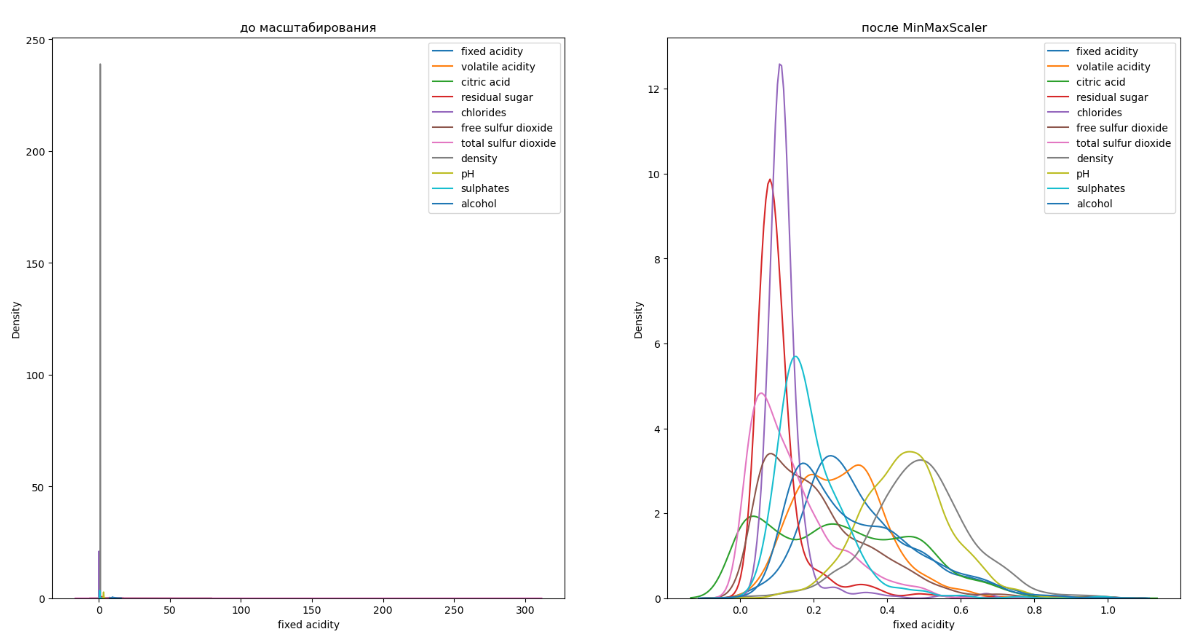
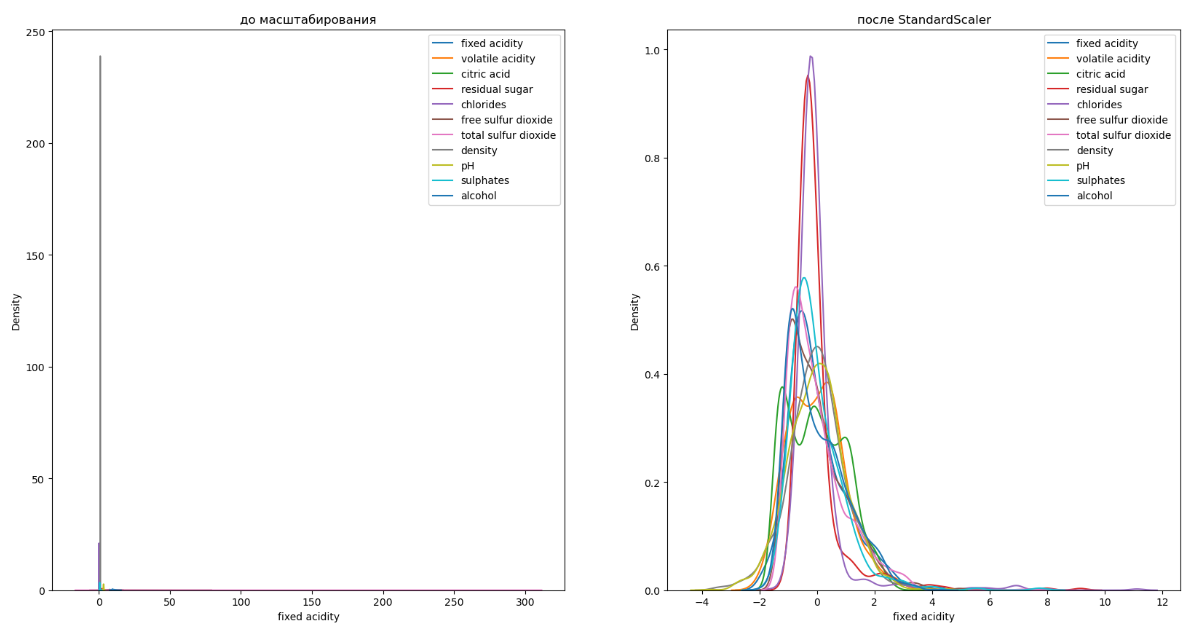
 

Рис. 2 – Кед-диаграммы трех методов масштабирования

1. Обработка выбросов для числовых признаков

Я использую IsolationForest для удаления провалов и замены провалов медианной заменой.



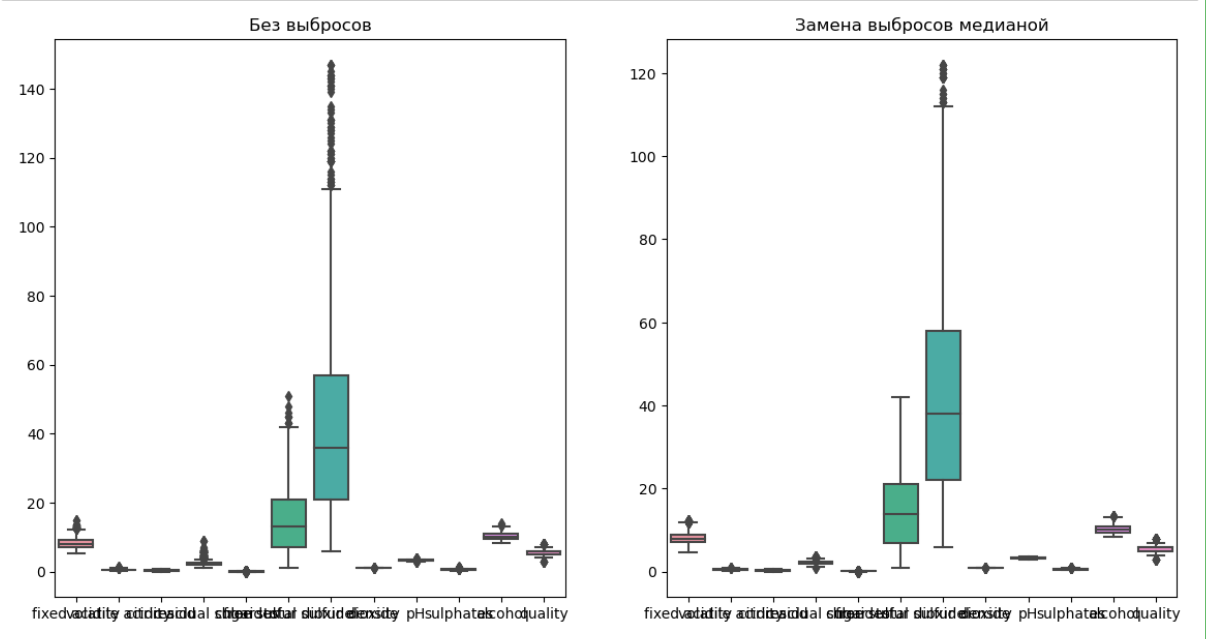
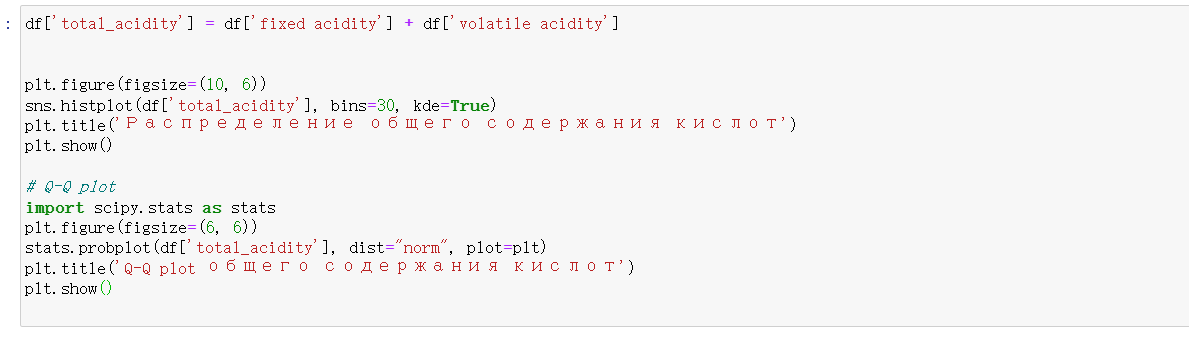


Рис.3– Сравнение распределений до и после обработки выбросов

1. Обработка по крайней мере одного нестандартного признака



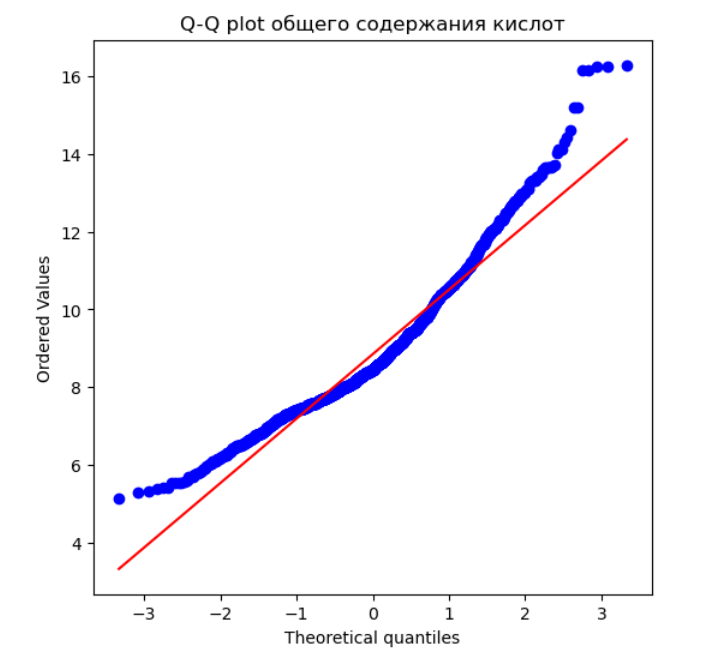
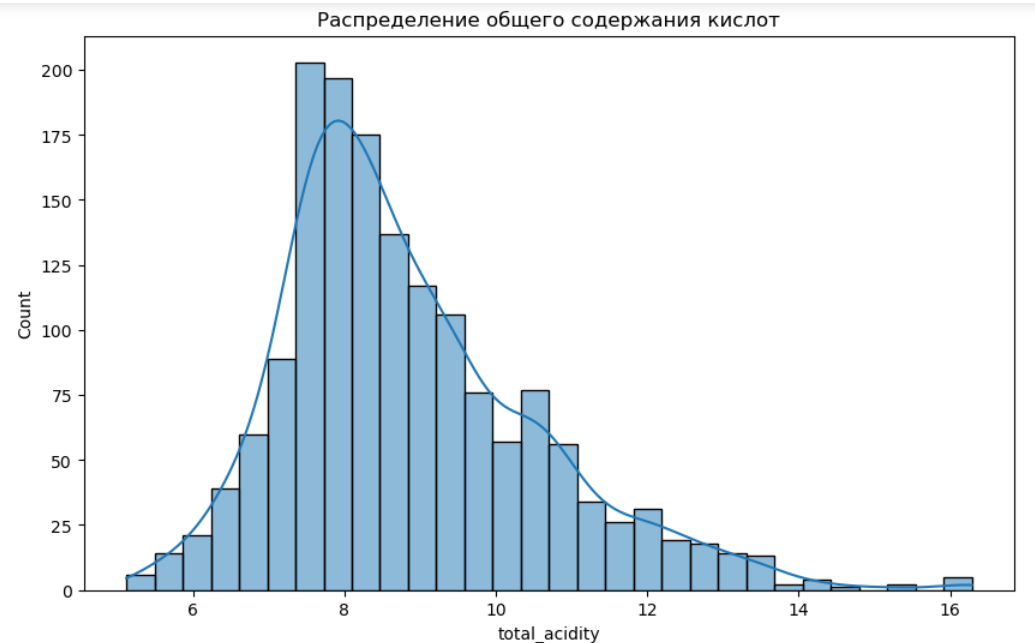


Рис.4– визуализируем его распределение и Q-Q plot.

1. Отбор признаков

Я использую методы SelectKBest (фильтрация), RFE (обертывание) и Lasso (встраивание).



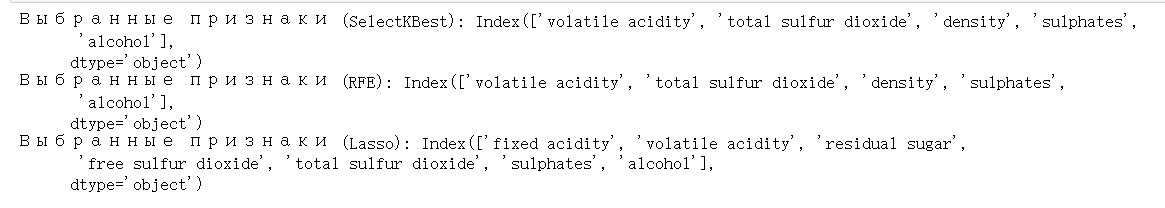


Рис.5–Отбор признаков