

**Московский авиационный институт
(Национальный исследовательский университет)**

Институт: «Информационные технологии и прикладная математика»

Кафедра: 806 «Вычислительная математика и программирование»

Дисциплина: «Искусственный интеллект»

Лабораторная работа № 0

Студент: Батова Е.Д.

Группа: М8О-301Б-19

Дата:

Оценка:

Москва, 2022

1. Постановка задачи

Постановка задачи о качестве производства вина. Проанализировать данные для набора данных “Red Wine Quality”

2. Описание программы

Никогда не хочется заслужить пищевое отравление, употребляя некачественные еду или напитки. Анализ показателей вина может помочь сделать выбор в пользу более качественной продукции. То есть нужно дать оценку для параметров. Оценка будет от 3 до 8.

3. Анализ данных

В датасете представлены следующие признаки:

1 - fixed acidity - параметр указывает на количество кислотности (винная, яблочная и янтарная) в вине.

2 - volatile acidity - параметр относится к присутствующим в вине кислотам, перегоняемым с водяным паром.

3 - citric acid - параметр указывает на содержание лимонной кислоты.

4 - residual sugar - параметр показывает количество углеводов, которое осталось после завершения процесса брожения

5 - chlorides - параметр указывает на содержание соли в вине.

6 - free sulfur dioxide - параметр показывает количество диоксида серы, который влияет на противогрибковые свойства.

7 - total sulfur dioxide - Параметр указывает на общее количество диоксида серы

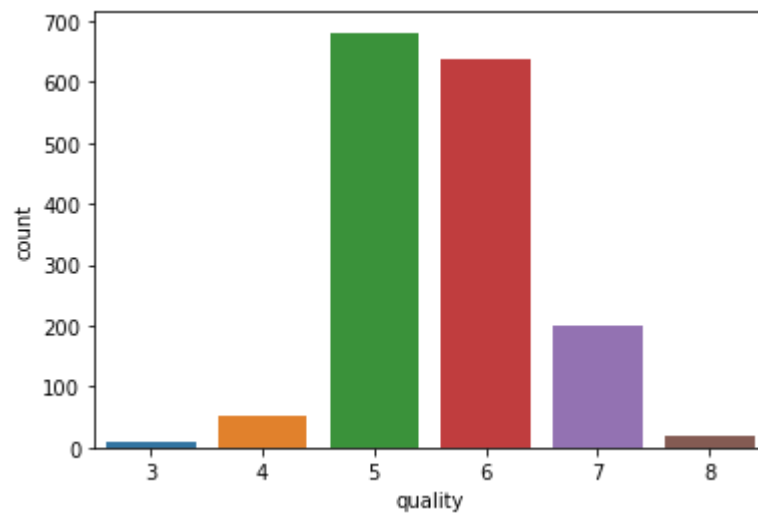
8 - density - параметр определяет плотность вина. Плотность вина зависит от содержания в нем спирта, остаточного сахара, глицерина.

9 - pH - параметр описывает уровень кислотности по шкале от 0 до 14.

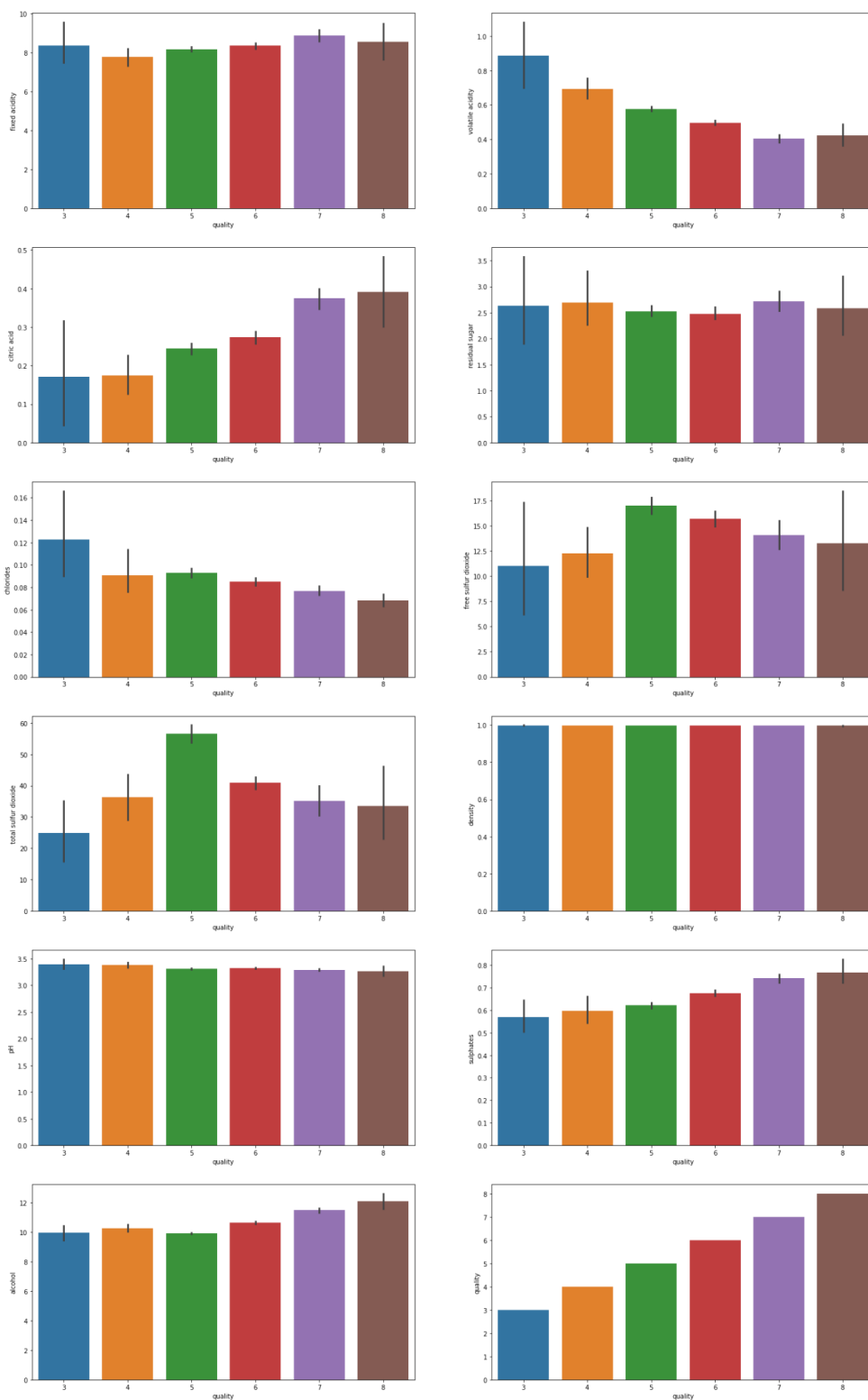
10 - sulphates - параметр описывает добавку, которая усиливает действия диоксида серы.

11 - alcohol - параметр указывает содержания алкоголя в вине.

12 - quality - сама оценка

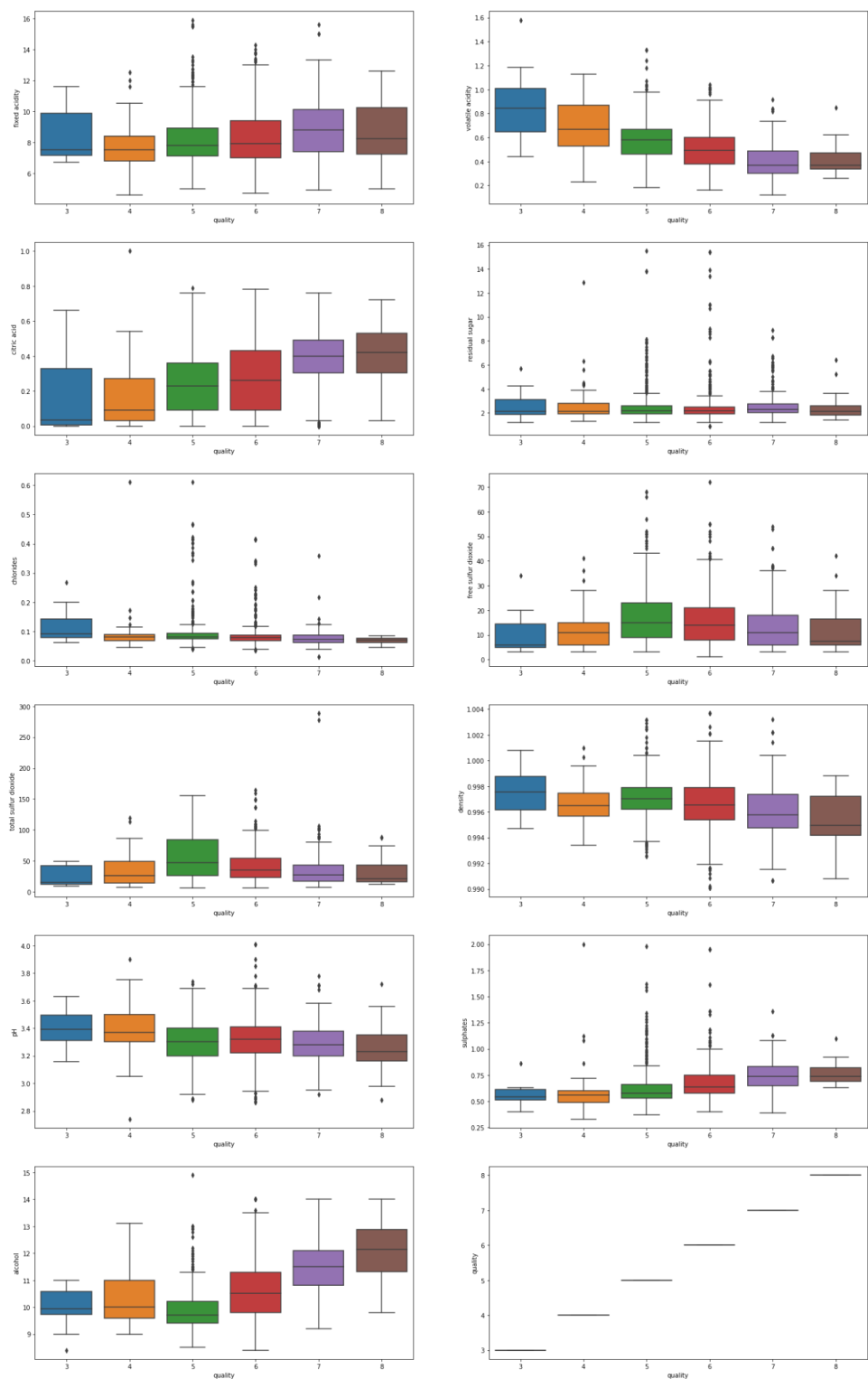


Из графика видно, что хорошего вина, как и совсем плохого, меньше, чем среднего качества. Получается, что данные несбалансированные.

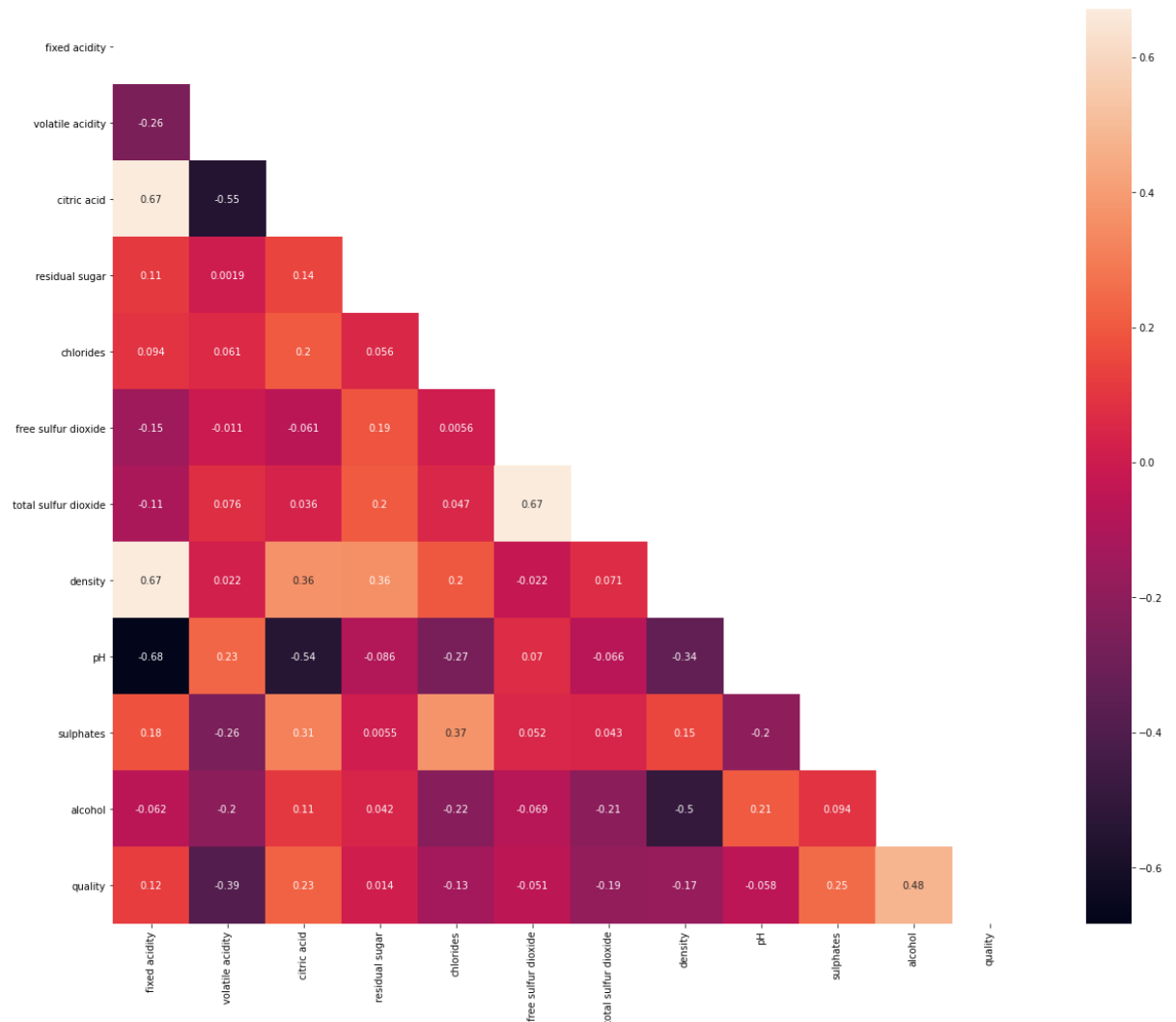


Данные графики показывают зависимость оценки вина от различных параметров. Как видно, не на всех параметрах хорошо видна зависимость, но некоторые из них отражают тенденцию. Например, fixed acidity, residual sugar, density, pH несильно влияют на качество вина. volatile acidity, chlorides видно, что при снижении этих

параметров качество увеличивается. Наоборот происходит с citric acid, sulphates, alcohol - при увеличении параметров увеличивается и качество. Погрешность некоторых параметров достаточно большая, нужно смотреть выбросы.



Выбросов оказалось достаточно много, что нужно будет также корректировать.



Можно заметить, что корреляция между параметрами не такая уж и сильная. Самая сильная корреляция равна 0.67. Так что какие-либо параметры трудно убрать.

4. Вывод

Перед началом обучения нужно будет скорректировать данные, так чтобы не возникло неприятностей в дальнейшем. Также нужно удалить или откорректировать выбросы. Чтобы уменьшить количество параметров, можно убрать наименее коррелированные.