**Определение качества вина**

**1. Общее представление**

**1.1. Введение**

Качество вина – это совокупность его свойств, влияющих на восприятие вина потребителем.

**1.2. Проблема**

Качество является совокупностью приятных вкусовых ощущений, непосредственно связанных с

химическим составом вина. В то же время само понятие “качества” является субъективным и зависит от предпочтений потребителя. Однако, в рамках данной задачи у нас имеются объективные химические показатели вина и оценки качества, сделанные, как предполагается, профессионалами, и которые тоже можно принять за объективный показатель в рамках данного исследования.

**1.3. Решение**

Анализируем имеющиеся в распоряжении данные характеристик вин, выделяем нужные признаки,

строим модель классификации качества вин и оцениваем ее качество. На выходе получаем

модель оценки качества вина, которая может быть использована для оценки других вин на основе их химического состава.

**2. Реализация**

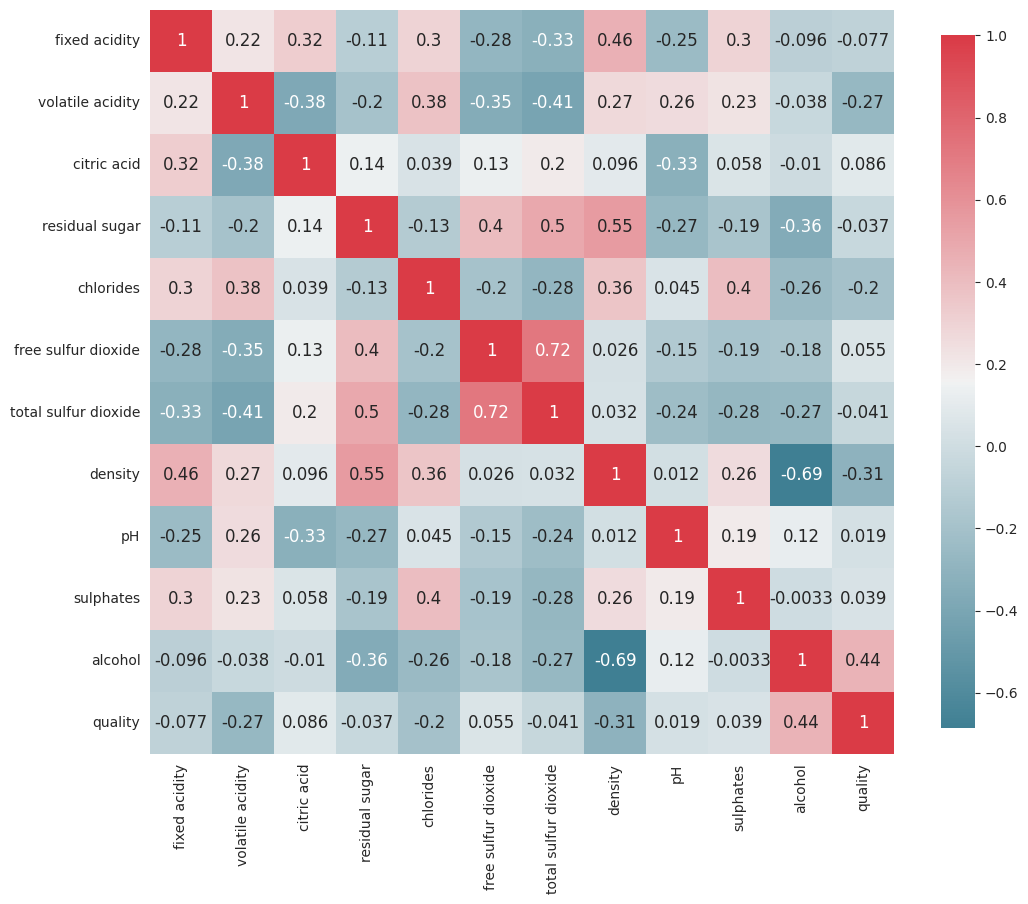
**2.1 Анализ данных и выделение нужных признаков**

Для анализа представлен датасет (<https://www.kaggle.com/rajyellow46/wine-quality>), состоящий из 11

характеристик около 6500 вин и оценок их качества профессиональными сомелье по шкале от 0 до 10.

Характеристики и их корреляция представлены на изображении ниже. Есть несколько признаков, скоррелированных с качеством вина, однако, корреляция не настолько велика, чтобы только лишь по ним определять качество.

Матрица корреляции:



Взаимная корреляция признаков не достаточно высока, чтобы исключить какие-то признаки.

Датасет не требует большого объема обработки: пропусков не много, явных выбросов по значениям нет. Требуется нормализация характеристик датасета перед построением модели.

**2.2. Построение модели и оценка качества**

Строим 2 варианта модели, чтобы можно было выявить среди них наиболее удачную: KNN и RandomForest.

Ниже представлена таблица оценки качества (score) обоих вариантов моделей на тренировочной и

тестовой выборках.

|  |  |  |
| --- | --- | --- |
| Модель | Train data | Test\_data |
| KNN | 0.67 | 0.56 |
| RandomForest | 0.83 | 0.62 |

**3. Заключение**

В результате проделанной работы была построена модель, способная на наборе характеристик вина

предсказывать его качество. Модель была представлена в 2-х вариантах. Лучше себя показал вариант RandomForest. Тем не менее, качество модели далеко от идеального.

При решении реальной рабочей задачи следует испробовать большее число моделей с различными вариантами гиперпараметров.