

## Kaggle

<https://www.kaggle.com/rajyellow46/wine-quality>

### Описание проблемы

Задача состоит в том, чтобы предсказать качество красного вина по шкале от 0 до 10 с учетом набора характеристик в качестве входных данных.

### Обоснование (бизнес-проблема)

Когда мы сможем определять на что влияет качество вина, мы сможем улучшить показатели производства и соответственно повысим лояльность к продукту.

### Особенности и обработка данных

Входными переменными для решения задачи у нас были следующие: фиксированная кислотность, прочая кислотность, лимонная кислота, остаточный сахар, хлориды, свободный диоксид серы, общий диоксид серы, плотность, pH, сульфаты, спирт.

Выходной переменной (на основе сенсорных данных) является качество (оценка от 0 до 10).

### Модель обучения

Для обучения модели мы использовали линейную регрессию.

Линейная регрессия - модель зависимости переменной  $x$  от одной или нескольких других переменных (факторов, регрессоров, независимых переменных) с линейной функцией зависимости.

### Вывод

Из результатов поняли: что при фиксированных всех других характеристиках увеличение содержания сульфатов на 1 единицу приведет к повышению качества вина на 0,8 и аналогично другим функциям.

При сохранении всех остальных функций увеличение летучей кислотности на 1 единицу приведет к снижению качества вина на 0,99, а также к другим характеристикам.

Таким образом, с помощью нескольких строк кода мы смогли построить модель линейной регрессии для прогнозирования качества вина с показателями RMSE 0,74 и 0,72 для тренировок и тестирования соответственно.