Предсказание качества вина

С.Федюшин s.fedyushin9211@gmail.com

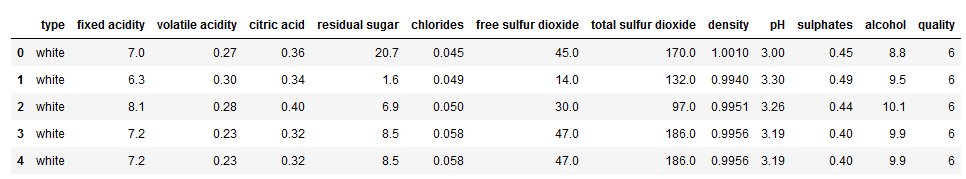
# 1 Введение

Разработана модель предсказания качества вина по его физико-химическим и сенсорным качествам.

# 2 Описание проблемы и данные

DataSet был загружен из репозитория машинного UCI.

Два набора данных связаны с красными и белыми вариантами португальского вина «Vinho Verde». Ссылка [Cortez et al., 2009]. Из-за конфиденциальности и логистических проблем доступны только физико-химические (входы) и сенсорные (выходные) переменные (например, нет данных о типах винограда, винодельческой марки, цены продажи вина и т. д.).



Входные переменные (основанные на физико-химических тестах):

type – тип вина (красное или белое);

fixed acidity – титрируемая кислотность;

volatile acidity – летучая кислотность;

citric acid – лимонная кислота;

residual sugar – сахар;

chlorides – хлориды;

free sulfur dioxide – свободный диоксид серы;

total sulfur dioxide – общий диоксид серы;

density – плотность;

pH – кислотность;

sulphates – сульфаты;

alcohol – объемная доля этилового спирта.

Выходная переменная (основанная на сенсорных данных):

quality – качество (балл в диапазоне от 1 до 10).

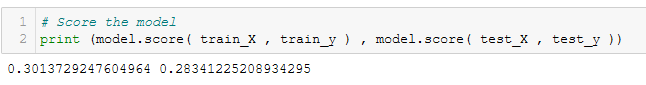
# 3 Признаки и предобработка данных

Пропуски данных были заполнены средними значениями, категориальная переменная type была закодирована через 1 и 0.

Проведено разбиение данных на обучающие и тестовые в пропорции 70/30.

# 4 Моделирование

4.1 В качестве отправной точки была выбрана линейная регрессия. Проведено обучение и валидация модели.



4.2 Точность на обучающей и тестовой выборках не отличается, значит модель не переобучилась.

4.3 Среднеквадратичное отклонение на тестовых данных составило 0,58.

# 5 Рекомендации по улучшению модели

Для улучшения результатов предлагаются следующие методы, связанные с этапами предобработки данных и постороения модели, а именно:

- выравнивание баланса классов;

- нормализация признаков;

- отсеивание строк с выбросами;

- использование других моделей как классификационных, так и регрессионных с подбором гиперпараметров.