

Отчет по прогнозированию качества вин

Что я сделал:

1. Проверил три варианта моделей:
 - Общая модель для всех вин: R^2 0.495, ошибка 0.448
 - Отдельно для белых: R^2 0.483, ошибка 0.463
 - Отдельно для красных: R^2 0.541, ошибка 0.440

Главные выводы:

1. Красные и белые вина лучше предсказывать разными моделями - они работают точнее, чем общая.
2. Причины:
 - У красных и белых вин разный химический состав
 - В данных мало редких примеров (очень плохих и очень хороших вин)

Что можно улучшить (если позволят условия):

1. **Поменять алгоритм**
Нейросети - не самый лучший выбор для таких табличных данных.
(по моему мнению)
Плюсы:
 - Быстрее настраиваются
 - Лучше работают с малыми данными
 - Проще понять, как модель принимает решения
2. **Доработать данные**
 - Добавить новые признаки
 - Увеличить количество примеров редких классов (вин с оценками 3,4 и 8,9)
 - Настроить веса, чтобы модель больше внимания уделяла редким случаям

Итог:

С текущими данными и нейросетью мы получили хороший результат. Но для реального применения лучше использовать другие алгоритмы (CatBoost/XGBoost) - они дадут большую точность с меньшими усилиями.