

Практическое задание №1.6.6

Для выполнения практического задания будем использовать файл «winequality-red.csv», содержащий результаты физико-химических тестов красного вина.

Источник: <https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009>.

Задача: построить модель случайного леса, определяющую хорошее вино или плохое на основе результатов физико-химических тестов.

Входные переменные (признаки):

- fixed acidity – фиксированная кислотность
- volatile acidity – летучая кислотность
- citric acid – содержание лимонной кислоты
- residual sugar – остаточный сахар
- chlorides – хлориды
- free sulfur dioxide – свободный диоксид серы
- total sulfur dioxide – общий диоксид серы
- density – плотность
- pH
- sulphates – сульфаты
- alcohol – спирт

Целевая переменная – quality (качество): 0 – плохое вино, 1 – хорошее вино.

1. Загрузите набор данных из файла «winequality-red.csv» в DataFrame.
2. Обучите модель случайного леса ([RandomForestClassifier](#)).
3. Оцените качество модели. Метрики для оценки качества выберите самостоятельно.
4. Оцените информативность признаков с помощью атрибута `feature_importances_`.
4. Поимпровизируйте с этими данными. Попробуйте построить другие модели (на ваш выбор). Также, можно попробовать изменить значения гиперпараметров моделей и посмотреть, как изменится результат.

