

Задача №4 – Регрессоры

Написать программу на Python, которая обучает три регрессионных модели, построенных на наборе с помощью трёх алгоритмов: линейный регрессор, полиномиальный регрессор и регрессор, основанный на случайном лесе.

- Выбрать признаки, используемые при обучении, и, если необходимо, выполнить их предобработку. Разделить выборку на обучающую и тестовую.
- В работе необходимо исследовать работу алгоритма случайный лес с разными значениями гиперпараметров и степенью полинома для полиномиальной регрессионной модели. Для модели *случайный лес* вывести значения важности признаков.
- Написать короткий отчет по работе, включив в него программу с комментариями, значения качества моделей (коэффициент детерминации, среднюю квадратичную и среднюю абсолютную ошибки). Выбрать наилучшую модель из полученных регрессоров.

Выберите свой вариант согласно последней цифре вашего номера в журнале:

- если последняя цифра 0 или 5: датасет – Лесные пожары (<https://archive.ics.uci.edu/ml/datasets/Forest+Fires>), предсказываемое значение – площадь пожара (Area);
- если последняя цифра 1 или 6: датасет – Качество вина (<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>) предсказываемое значение – качество (Quality), для датасета с красным вином, winequality-red.csv;
- если последняя цифра 2 или 7: датасет – Качество вина (<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>) предсказываемое значение – качество (Quality), для датасета с белым вином, winequality-white.csv;
- если последняя цифра 3 или 8: датасет – Аренда велосипедов (<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>), предсказываемое значение – количество аренд велосипедов в сутки (Area), датасет day.csv;
- если последняя цифра 4 или 9: датасет – Аренда велосипедов (<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>), предсказываемое значение – количество аренд велосипедов в час (Area), датасет hour.csv;