Задание №3 – Регрессия

Написать программу на Python, которая обучает четыре регрессионных модели, построенных на наборе с помощью четырёх алгоритмов: линейный регрессор, полиномиальный регрессор, регрессор, основанный на случайном лесе и один из следующих инструментов: Gaussian Process Regression, Support Vector Regression, Gradient Boosting Regressor или AdaBoost Regressor.
Выбрать признаки, использующиеся при обучении, и, если необходимо, выполнить их предобработку. Разделить выборку на обучающую и тестовую.
В работе необходимо исследовать работу алгоритмов с разными значениями гиперпараметров.
Для моделей на основе деревьев вывести значения важности признаков.
Выбрать наилучшую модель из полученных регрессоров.
Сохранить лучшую модель (pickle). Спроектировать и реализовать приложение (настольное, Web- или чат-бот) на Python, осуществляющее загрузку модели, проверку корректности ввода данных, требуемых для регрессионной модели. Для реализации программы можно использовать любой инструмент (Flask, Django, Tkinter и т.д.)
Написать короткий отчет по работе, включив в него программы с комментариями, значения качества моделей (коэффициент детерминации, среднюю квадратичную и среднюю абсолютную ошибки).

Для своего варианта необходимо посмотреть последнюю цифру номера своей зачетной книжки (или студенческого билета) и выполнить следующие корректировки:

- если последняя цифра 0 или 5: выборка – Лесные пожары (https://archive.ics.uci.edu/ml/datasets/Forest+Fires), предсказываемое значение – площадь пожара (Area);
- если последняя цифра 1 или 6: выборка – Качество вина (https://archive.ics.uci.edu/ml/datasets/Wine+Quality) предсказываемое значение – качество (Quality), файл winequality-red.csv;
- если последняя цифра 2 или 7: выборка – Качество вина (https://archive.ics.uci.edu/ml/datasets/Wine+Quality) предсказываемое значение – качество (Quality), файл winequality-white.csv;
- если последняя цифра 3 или 8: выборка – Аренда велосипедов (https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset), предсказываемое значение – количество аренд велосипедов в сутки (Area), файл day.csv;

• если последняя цифра 4 или 9: выборка – Аренда велосипедов (https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset), предсказываемое значение – количество аренд велосипедов в час (Area), файл hour.csv;