

Отчет по заданию  
«Ансамбли алгоритмов. Веб-сервер. Композиции  
алгоритмов для решения задачи регрессии.»  
для курса «Практикум на ЭВМ 2021/2022» для  
бакалавров 3 курса кафедры ММП ВМК МГУ

Фролов А., 317 группа, ВМК МГУ

7 декабря 2021

## 1 Предобработка данных

Для проведения экспериментов выбрана задача предсказания стоимости квартир. Датасет состоит из 21613 записей. Для предобработки признаки были разделены на три группы:

- численные (и ординальные): bedrooms, bathrooms, sqft\_living, sqft\_lot, floors, condition, grade, sqft\_above, sqft\_basement, yr\_built, yr\_renovated, sqft\_living15, sqft\_lot15
- бинарные: waterfront
- категориальные: view, zipcode, zone

Был добавлен новый признак zone, показывающий находится ли дом в аномальном участке (цена отклонилась от средней больше чем на  $3\sigma$ ). Для этого местность была поделена на 400 зон. В столбце yr\_renovated нули были заменены на значения из столбца yr\_built.

Численные признаки были нормализованы, категориальные были преобразованы с помощью One-Hot кодирования, бинарные не изменялись. Колонки id, date, lat, long были удалены. После предобработки размер признакового пространства увеличился до 92 признаков.

## 2 Случайный лес

## 3 Градиентный бустинг