

# Дипломная работа

По теме:

**Разработка модели для  
предсказания цен на  
недвижимость**

Выполнил: Досаев Савелий

# Актуальность задачи

Актуальность темы дипломной работы обусловлена тем, что на сегодняшний день появляется все большая необходимость в оценке стоимости недвижимости

## Цель работы

Предсказание рыночной стоимости объекта недвижимости с использованием машинного обучения.

## Оценка стоимости имущества необходима:

- ▶ при операциях купли-продажи или сдачи в аренду;
- ▶ при акционировании предприятий;
- ▶ при кадастровой оценке для целей налогообложения объектов недвижимости;
- ▶ для страхования объектов недвижимости;
- ▶ при кредитовании объектов недвижимости;
- ▶ при разработке инвестиционных проектов и привлечения инвесторов;
- ▶ при ликвидации объектов недвижимости;
- ▶ при исполнении прав наследования, судебного приговора;

# Для достижения поставленной цели необходимо решить следующие задачи:

- ▶ Обработать данные и привести их к числовому формату
- ▶ Очистить данные от вбросов
- ▶ Проанализировать важность различных признаков
- ▶ Отобрать признаки для обучения
- ▶ Выбрать алгоритм машинного обучения и обучить модель
- ▶ Оценить полученный результат

# Оценка полученного результата

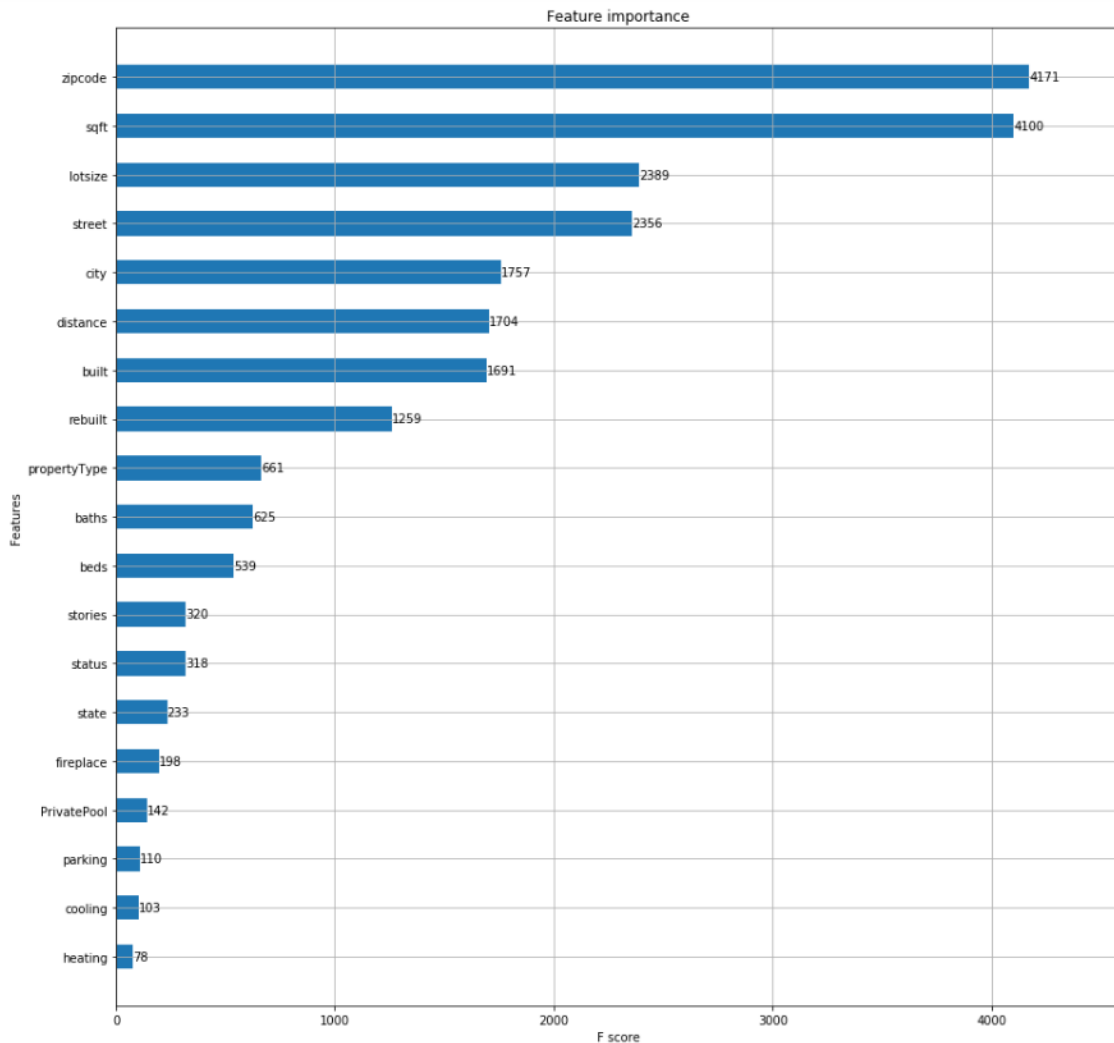
## ► Основная метрика:

- Mape(Mean absolute percentage error)
- Средний процент ошибки: принимает значения от 0 до  $\infty$

## Дополнительные метрики:

- R2(Коэффициент детерминации)
- Доля необъяснённой дисперсии: принимает значения от  $-\infty$  до 1
- Rmse(Root-mean-square deviation)
- Корень разности между предсказанным и настоящим значениями. Принимает значения от 0 до  $\infty$

# Факторы, влияющие на СТОИМОСТЬ НЕДВИЖИМОСТИ



Адрес и площадь  
имеют наибольший  
вклад в СТОИМОСТЬ

Heating, cooling,  
parking, state,  
privatePool и  
fireplace  
дают незначитель  
ный результат

# Недвижимость для прогнозирования

В ходе работы я обучаю модель на очищенных данных  
Из 360000 домов для обучения пошло только 200000 из них

- ▶ Лучшим образом для предсказания покажут себя:
- ▶ Квартиры с площадью от 20 кв. м.
- ▶ Любые дома до 10 этажей и с площадью до 650 кв. м.
- ▶ Участки до 4 кв. км.
  
- ▶ Модель плохо работает на:
- ▶ Больших участках или фермах
- ▶ Дешевой недвижимости(<40000\$)
- ▶ Аномально больших домах(кол-во ванн, этажей, спален>10, цена>3500000\$)

# Предсказания по разным признакам

Точность с использованием всех данных

|   | Model         | r2_train | r2_test | mape_train | mape_test | rmse_train  | rmse_test   |
|---|---------------|----------|---------|------------|-----------|-------------|-------------|
| 0 | Random Forest | 63.61    | 61.70   | 47.39      | 47.73     | 27980948.58 | 28536709.18 |
| 1 | XGB           | 97.04    | 82.81   | 15.71      | 24.20     | 7983087.19  | 19117250.21 |
| 2 | LGBM          | 96.72    | 84.35   | 4.92       | 22.40     | 8395104.36  | 18238378.22 |

Точность с использованием только адреса и площади

|   | Model         | r2_train | r2_test | mape_train | mape_test | rmse_train  | rmse_test   |
|---|---------------|----------|---------|------------|-----------|-------------|-------------|
| 0 | Random Forest | 63.54    | 62.09   | 47.32      | 47.62     | 28009618.79 | 28392310.92 |
| 1 | XGB           | 95.00    | 82.03   | 19.39      | 26.22     | 10368715.39 | 19544315.79 |
| 2 | LGBM          | 96.55    | 83.46   | 5.96       | 24.15     | 8620859.69  | 18753912.22 |

Точность падает только на 1.75% по сравнению с предсказаниями по всем данным



# Итог работы

- ▶ Обученная модель для предсказания цен на недвижимость
  - ▶ Точность модели - ~78%
  - ▶ Среднее отклонение от действительной цены - ~90000
- ▶ Прототип - работающая модель
  - ▶ Принимает в себя такие данные как:
    - ▶ Адрес и площадь
    - ▶ год постройки и реновации
    - ▶ Спальни, этажи и ванные комнаты
    - ▶ Тип и статус строения

По времени прототип работает:

6 секунд для предсказания 1 дома

10 секунд для предсказания 10000 домов

# https://github.com/Brilliance1512/dataproject

The screenshot shows the GitHub repository page for `Brilliance1512/dataproject`. The page layout includes a header with navigation links, a repository overview section with statistics, a file list, and a footer with copyright information.

**Repository Overview:**

- Repository: `Brilliance1512 / dataproject`
- Unwatch: 1, Star: 0, Fork: 0
- Code (selected), Issues: 0, Pull requests: 0, Actions, Projects: 0, Wiki, Security: 0, Insights, Settings
- No description, website, or topics provided. [Edit](#)
- Manage topics
- 47 commits, 1 branch, 0 packages, 0 releases, 1 contributor
- Branch: master, New pull request
- Create new file, Upload files, Find file, Clone or download

**File List:**

| File Name            | Commit Message              | Time Ago    |
|----------------------|-----------------------------|-------------|
| Описание проекта.txt | Update Описание проекта.txt | 24 days ago |
| Проект.ipynb         | Add files via upload        | 7 days ago  |
| Прототип.ipynb       | Add files via upload        | 2 hours ago |

**Footer:**

Help people interested in this repository understand your project by adding a README. [Add a README](#)

© 2020 GitHub, Inc. [Terms](#) [Privacy](#) [Security](#) [Status](#) [Help](#) [Contact GitHub](#) [Pricing](#) [API](#) [Training](#) [Blog](#) [About](#)