

NskTeam



Сервис расчета рыночной стоимости
жилой недвижимости города Москвы
(задача №6)

Команда NskTeam

Роль:

- Написание кода (python): Front, Back, ML
- Идея и реализация алгоритма автоподгрузки объявлений ("парсинг")
- Data Science
- Контейнеризация и перенос на SberCloud
- Документация, презентация
- Генерация идей
- Поиск и обработка данных



Капитан

Александра
Плотникова

+7 913 ***

sasa2111@mail.ru

ТГ: @Alexandra_Plo



Участник

Надежда
Белова

+7 927 ***

***@outlook.com

ТГ: @***

Роль:

- Тестирование
- Генерация идей
- Поиск данных
- Поиск информации о различных методах парсинга

Схема сервиса (MVP)

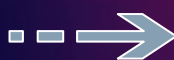
Яндекс. Недвижимость



Подписки на рассылку
новых объявлений



26 настроенных
фильтров = 26 рассылок.
Частота = max



Спец. почта на mail.ru



Почтовый фильтр
складывает все
письма в
определенную папку



Docker container



Модуль расчета
рыночной цены
эталона по ТЗ



Модуль подбора
аналогов



Модуль расчета
рыночной цены
всего пула



ML-модель
альтернативного
расчета цены



Архив
оценок



Модуль
автоподгрузки и
расшифровки
новых объявлений



Архив
объявлений

User web-interface

Сервис

Расчет рыночной
цены пула (2 оценки)

Запуск автоподгрузки
новых объявлений

Скачивание всех
оценок

Скачивание всех
объявлений

Скачивание шаблона
файла для загрузки



Характеристики интерфейса:

- Интерактивные карты
- Интерактивные таблицы
- Пошаговость основного алгоритма
- Минималистичный дизайн
- Многостраничность
- Адаптивность для мобильных устройств

Найдено 6 аналогов:

	segment	address	repair	balcon	rooms	st_metro	metr
173	вторичка	Москва, Кастанаевская улица, 45К2	евроремонт	balcon	1.0000	Славянский бульвар	
293	вторичка	Москва, Кутузовский проспект, 82	евроремонт	balcon	2.0000	Кунцевская	
333	вторичка	Москва, Давыдовская улица, 18	евроремонт	balcon	2.0000	Давыдовское	
628	новострой	Москва, Кастанаевская улица, 66	н/д	н/д	5.0000	Кунцевская	
1063	вторичка	Москва, улица Ивана Франко, 6	требуется ремонт	balcon	0.0000	Кунцевская	
1328	новострой	Москва, улица Ивана Франко, 6	евроремонт	н/д	4.0000	Кунцевская	

На карте изображены аналоги и эталон. Более яркий цвет означает, что в этой точке больше объектов.

ШАГ 2. Поиск аналогов для эталона в указанном радиусе

Укажите фильтры для поиска аналогов.

Выберите радиус поиска, км:

Допустимое отклонение по площади, м2:

Подобрать аналоги

ШАГ 3. Расчет удельной стоимости

Настроить параметры расчета:



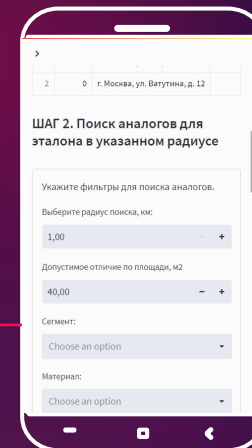
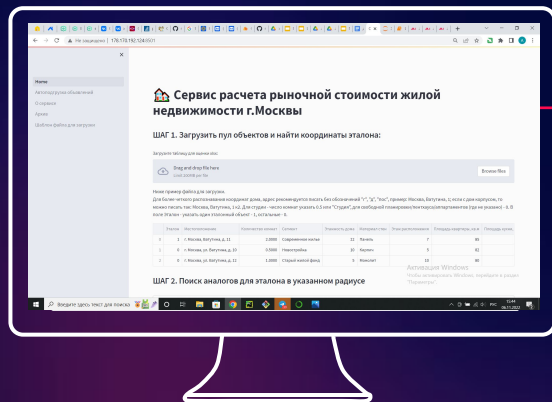
Интерфейс

Десктопная версия

Разделы сервиса:

- Home
- Автоподгрузка объявлений
- О сервисе
- Архив
- Шаблон файла для загрузки

Мобильная версия



Автоподгрузка новых объявлений

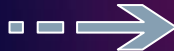
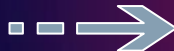
Яндекс. Недвижимость



Подписки на рассылку
новых объявлений



26 настроенных
фильтров = 26 рассылок.
Частота = max



Спец. почта на mail.ru



Почтовый фильтр
складывает все
письма в
определенную папку



Модуль автоподгрузки объявлений

- Запускается по запросу пользователя (кнопка)
- Смотрит непрочитанные письма
- Открывает их (отмеч. прочитанными)
- В каждом письме 1-5 объявлений - вытаскивает данные
- Расшифровывает данные о квартирах
- Подкачивает координаты по адресам (широта, долгота)
- Добавляет данные в общую таблицу объявлений



Архив
объявлений



Модули оценки пула

Оценка по ТЗ

1. Загрузка файла
2. Поиск координат эталона и вывод на карте
3. Фильтры
4. Подбор аналогов и вывод на карте
5. Корректировка фильтров/удаление части аналогов
6. Расчет корректировок для аналогов и стоимости эталона по ТЗ
7. Расчет стоимостей для пула по ТЗ

Оценка ИИ: CatboostRegressor

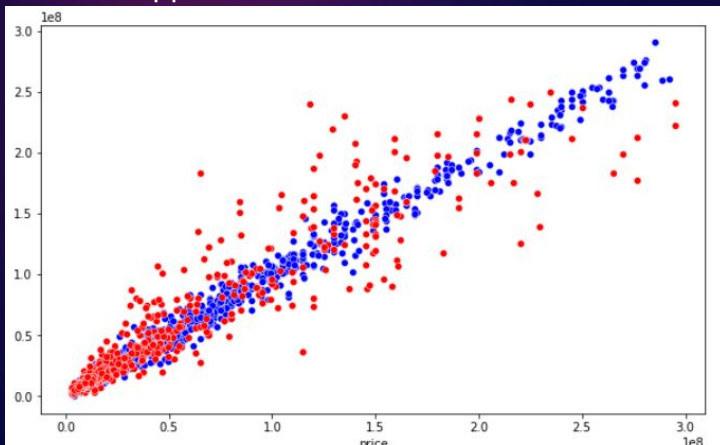
8. Адаптация данных для модели
9. Оценка стоимостей пула моделью

Таблица с 2-мя
оценками пула

ML-модель. Catboost.

1. Качество модели:
 - a. r^2 -score = 86% на тестовой выборке,
 - b. 98% на обучающей выборке.
 - c. см.графики
2. С чем сравнивалась: LinearRegression, LASSO, RidgeRegression (качество сильно хуже)
3. Модель обучена на:
 - a. более 2300 объявлений стоимостью до 300 млн в г.Мск,
 - b. период сбора данных с 24.10.2022 г. по 04.11.2022 г.

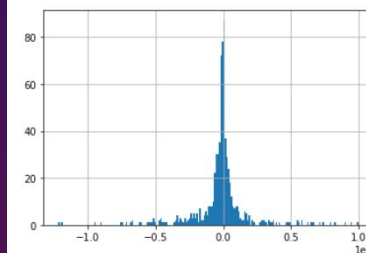
Гомоскедастичность



Дисперсия остатков catboost
намного стабильнее, чем у
других алгоритмов.
Распр. остатков близко к
нормальному.

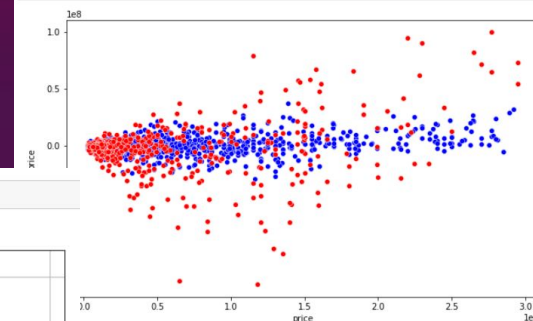
```
residual_test.hist(bins = 200)
```

<AxesSubplot:>



```
# посмотрим распределение остатков
residual_train = y_train - cat.predict(x_train)
residual_test = y_test - cat.predict(x_test)
figure = plt.subplots(figsize = (10,6))
sns.scatterplot(x=y_train, y=residual_train, color = 'blue')
sns.scatterplot(x=y_test, y=residual_test, color = 'red')

plt.show()
```



Язык программирования:	Python
Интерфейс (фронт):	Библиотека Streamlit: все элементы интерфейса + адаптив
Контейнеризация:	Docker (Прототип развернут на SberCloud)
Автоподгрузка объявлений:	Специально заведенная для проекта почта на mail.ru, + настроенные специальным образом подписки на рассылки объявлений на сайте Яндекс.недвижимость, + python: получение данных с почты и расшифровка - библиотеки imaplib, email, base64, dill, pandas, time, geopy
ML-модель:	CatboostRegressor
Архив и база объявлений:	Excel, Pickle (dill)
Использованные библиотеки Python:	pandas, datetime, time, imaplib, email, base64, streamlit, geopy, catboost, dill

Что сейчас в MVP	Как можно доработать
Сервис рассчитан на 1 пользователя, нет логинов и паролей	Мультипользовательская версия с правами доступа
Данные хранятся в нескольких таблицах xlsx, pkl	Полноценная БД
В сервисе нет модуля переобучения модели (не успела). Нужно переобучать вручную и загружать в сервис.	Добавить в сервис модуль с пайплайном переобучения модели, запускаемый пользователем или по расписанию.
База объявлений только пополняется	Добавить модуль чистки базы объявлений от слишком старых объявлений (более полугода, например)
Подгрузка новых объявлений происходит по нажатию кнопки.	Настроить автозапуск по расписанию (ночью например)
Почта и рассылки привязаны к личным телефонам капитана команды	Передать почту и аккаунты заказчику (заменить телефон / пересоздать и заново настроить аккаунты и подвязать к ним парсер)
ML-модель может быть улучшена, было очень ограниченное время на моделирование.	Более детальная работа с данными, улучшение качества модели, поиск доп. данных и фичей для обучения модели.

Home
Автоподгрузка объявлений
О сервисе
Архив
Шаблон файла для загрузки

Сервис расчета рыночной стоимости жилой недвижимости г.Москвы

ШАГ 1. Загрузить пул объектов и найти координаты эталона:

Загрузите таблицу для оценки xls:

Drag and drop file here
Limit 200MB per file

Browse files

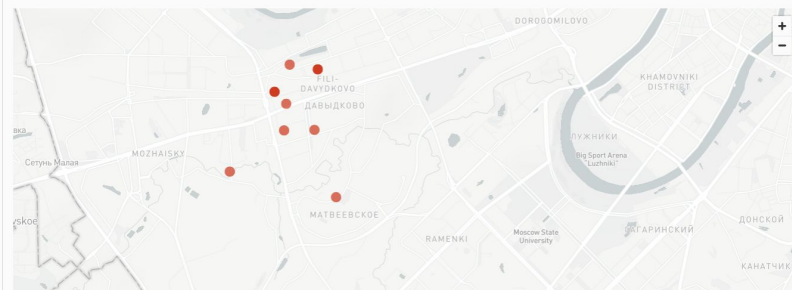
Ниже пример файла для загрузки.

Для более четкого распознавания координат дома, адрес рекомендуется писать без обозначений "г.", "д.", "пос", пример: Москва, Ватутина, 1; если с дом корпусом, то можно писать так: Москва, Ватутина, 1 к2. Для студии - число комнат указать 0.5 или "Студия", для свободной планировки/пентхауса/апартаментов (где не указано) - 0. В поле Эталон - указать один эталонный объект - 1, остальные - 0.

	Эталон	Местоположение	Количество комнат	Сегмент	Этажность дома	Материал стен	Этаж расположения	Площадь квартиры, кв.м	Площадь кухни,
0	1	г. Москва, Ватутина, д. 11	2.0000	Современное жилье	22	Панель	7	85	
1	0	г. Москва, ул. Ватутина, д. 10	0.5000	Новостройка	10	Кирпич	5	82	
2	0	г. Москва, ул. Ватутина, д. 12	1.0000	Старый жилой фонд	5	Монолит	10	80	

2603	вторичка	Москва, Некисная улица, Бк1	евроремонт	балкон	3.0000	Давыдовое	21.0000	105.5000	21.1000	18.0000	25.0000	монолит
2670	вторичка	Москва, Кастаневская улица, 45к2	евроремонт	балкон	1.0000	Славянский Бульвар	9.0000	50.8000	10.1600	14.0000	20.0000	панель

На карте изображены аналоги и эталон. Более яркий цвет означает, что в этой точке больше объектов.



Автоподгрузка объявлений для подбора аналогов

Здесь можно запустить работу модуля автоподгрузки последних объявлений о продаже квартир в систему.

Во время работы данного модуля не рекомендуется что-то еще делать в системе.

Если запускать автоподгрузку ежедневно, то она будет занимать не более нескольких минут. За одни сутки в обычно собирается несколько десятков объявлений.

Последнее обновление данных об объявлениях в системе: 2022-11-05 15:07:11.088002

Подгрузить новые объявления в систему

Процесс загрузки данных... может занять несколько минут.

Running mail_preparation(...)

Последнее обновление данных об объявлениях в системе: 2022-11-05

Подгрузить новые объявления в систему

Процесс загрузки данных... может занять несколько минут.

Подгрузка новых данных заняла 0.7274683117866516 мин.

Загружено 75 новых объявлений. Всего объявлений в базе: 2673.

Вы можете выкачать текущую базу объявлений в виде xls-файла

Скачать архив объявлений

Оценки пула по алгоритму и с помощью ИИ:

	Местоположение	Площадь квартиры, кв.м	price	ML_predict_price
0	г. Москва, Ватутина, д. 11	85	30,020,929.1446	29,211,723.5293
1	г. Москва, ул. Ватутина, д. 11	82	27,862,566.9395	23,729,503.8213
2	г. Москва, ул. Ватутина, д. 11	80	28,790,992.1361	33,079,005.7364

Скачать данные об оценке

Последнее обновление данных об объявлениях в системе: 2022-11-06 10:40:59.446965

Спасибо за внимание!