Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости России

Студент группы 11-922
Агниашвили Д.В.
Научный Руководитель:
к. т. н., доцент кафедры
программной инженерии
Кугуракова В.В.

Казанский федеральный университет
ИНСТИТУТ информационных технологий и интеллектуальных систем

Цель и задачи

Целью данной работы является проектирование и разработка веб-платформы, в том числе, методики оценки стоимости недвижимости с использованием нейросетевых технологий, объединяющую в себе несколько инструментов прогнозирования рынка.

Задачи работы:

1. **Анализ имеющихся наработок в данной области и возможность их использования в работе;**
2. **Отбор факторов, оказывающих влияние на рыночную стоимость квартир;**
3. **Подготовка обучающей выборки для нейронной сети;**
4. **Определение оптимального типа и характеристик нейронной сети, а так же метода ее обучения;**
5. **Проектирование и разработка инструмента прогнозирования оценки стоимости недвижимости.**

Актуальность

Научная

1. Более 2 тысяч публикаций на Web of Science по запросу «Realty prediction»
2. 100+ публикаций с уточнением «Deep learning»
3. 400+ публикаций с уточнением «machine learning»
4. Рост популярности направления Deep learning в данной сфере.

Жизненная

1. Доступ к реальной оценки недвижимости
2. Модернизация документооборота
3. Экономия времени риелтора и участников купли-продажи
4. Снижение экономических рисков
5. «Хайп» на новых разработках
6. «Уберизация»
7. Оптимизация процесса инвестирования в объекты недвижимости

Преимущества

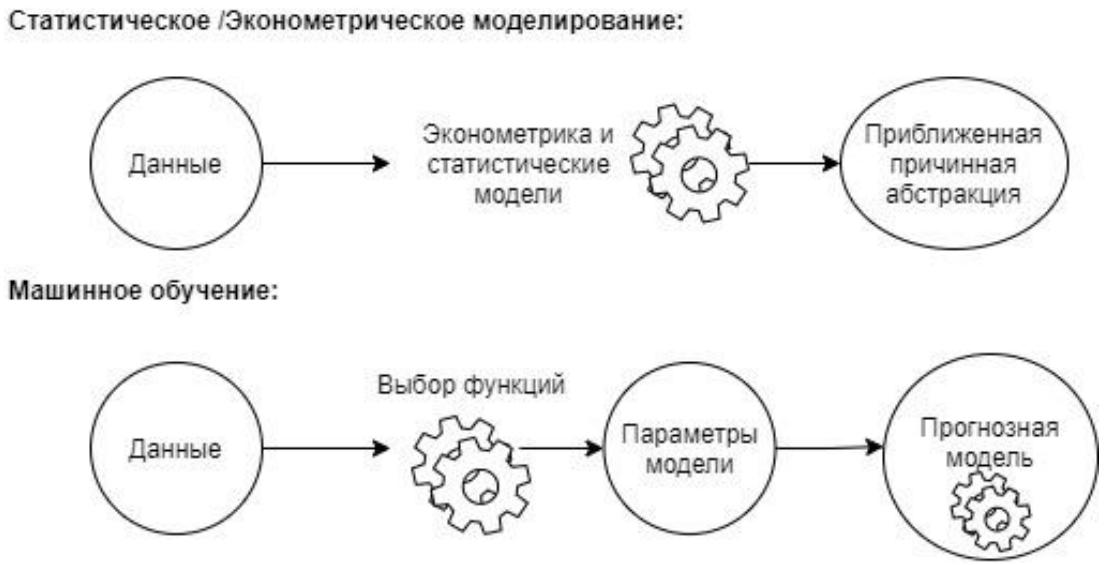
1. Решение проблемы отсутствия простых и наглядных сервисов прогнозирования и анализа недвижимости.
2. Помощь обывателям в инвестировании в недвижимость.
3. Систематизация функции оценки и прогнозирования стоимости жилых объектов на одной платформе.
4. Выявление и учет новых влияющих факторов на ценообразование жилых объектов.
5. Учет визуальных данных (фотографии с объявлений).
6. Сокращение лишних человеко-часов в условиях оценки объектов недвижимости.
7. Решение проблемы человеческого фактора.

Рынок недвижимости РФ

Рынок недвижимости Российской Федерации подвержен многочисленным колебаниям цен из-за существующих корреляций со многими переменными, некоторые из которых невозможно контролировать или даже могут быть неизвестны

Проблемы рынка недвижимости РФ:

1) Закрытость информации.
2) Молодой рынок. Сформирован в начале 1990-ых.
3) На сервисах Avito, Cian.Ru и других существуют только предлагаемая стоимость объекта, но нет реальных цен.
4) Разная макроэкономическая модель поведения стоимости.
5) Правила оценки недвижимости прописаны законодательством РФ и не может включать в себя «черный ящик».
6) Необходимо постоянно обновлять информацию.



Различия в методологиях моделирования

АНАЛОГИ

1) Cian.ru

Классификатор k-ближайших соседей. Сравнивает квартиру с местонахождением покупателя со множеством похожих объявлений из базы данных Cian за последние 5 лет

2) IRN.ru

Непонятно как работает алгоритм. Только по Москве и МО.

3) Domofond.ru

Основан на ценах объявлений, размещенных на Domofond.ru и Avito.ru. Средняя цена рассчитывается с учетом квартиры и квартир по соседству. Чем ближе будет квартира, тем больше будет ее вес в формуле

4) Ocenchik.ru

Особо не разрекламирован. Непонятно как работает алгоритм. Только по Москве и МО.

Используемые технологии

06

1. Языки программирования Python/C/SQL
2. СУБД PostgreSQL
3. Flask Framework
4. Grafana

Веб-приложение для аналитики и интерактивной визуализации с открытым исходным кодом

5. Optuna Framework

Фреймворк для автоматической оптимизации гиперпараметров

6. Scikit learn

Библиотека машинного обучения

Рабочий сервер

Домашний сервер

Используемые API

1. ADS-API.ru

Парсер недвижимости России. Собирается с avito.ru, irr.ru, realty.yandex.ru, cian.ru, sob.ru, youla.io, n1.ru и moyareklama.ru

2. DaData.ru

Автоматически проверяет, исправляет и геокодирует адреса

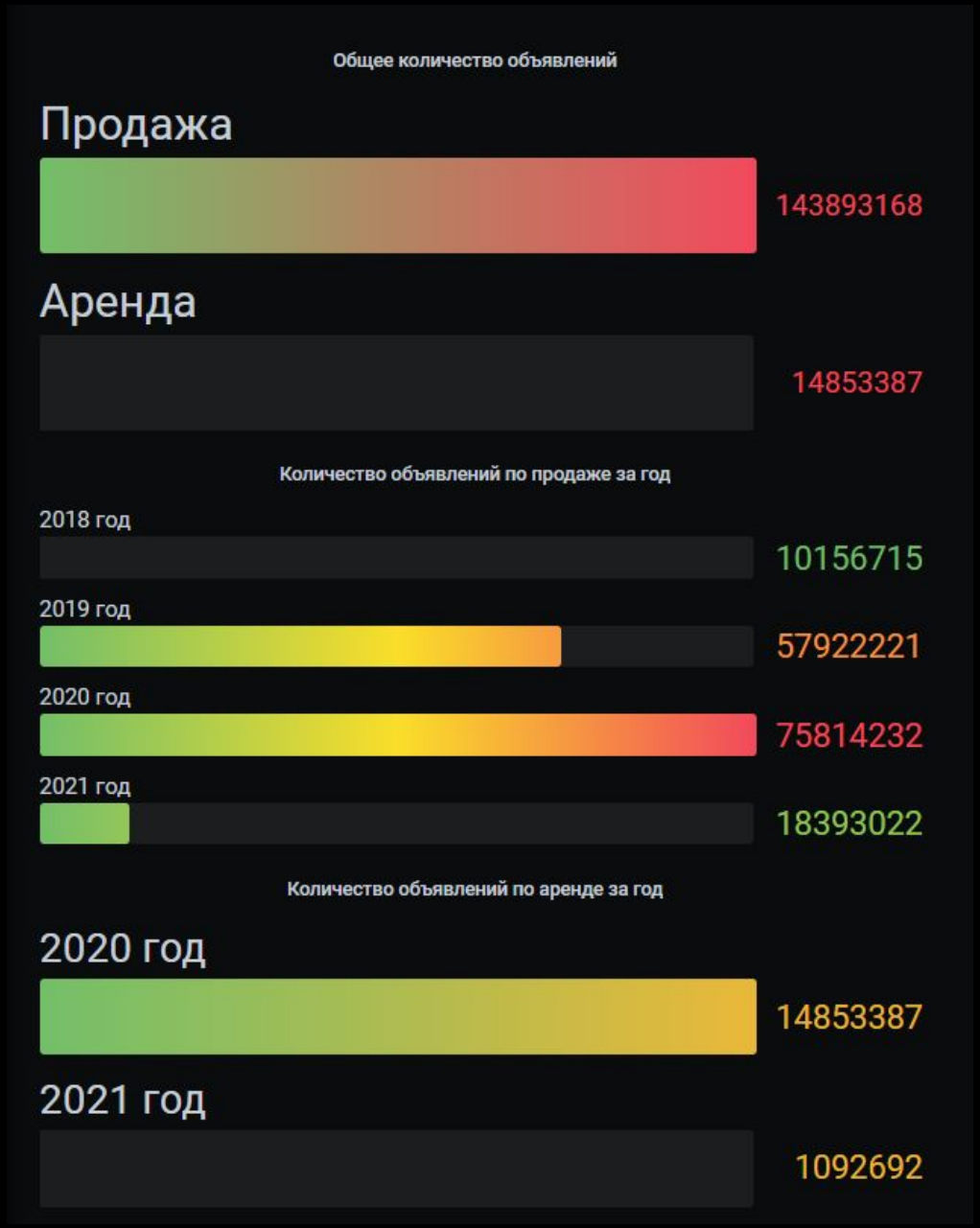
3. «Реформа ЖКХ»

Уникальная информационная система, объединившая оперативную информацию о том, что происходит с Вашим собственным жильем в частности и отраслью ЖКХ в целом

ДАТАСЕТ

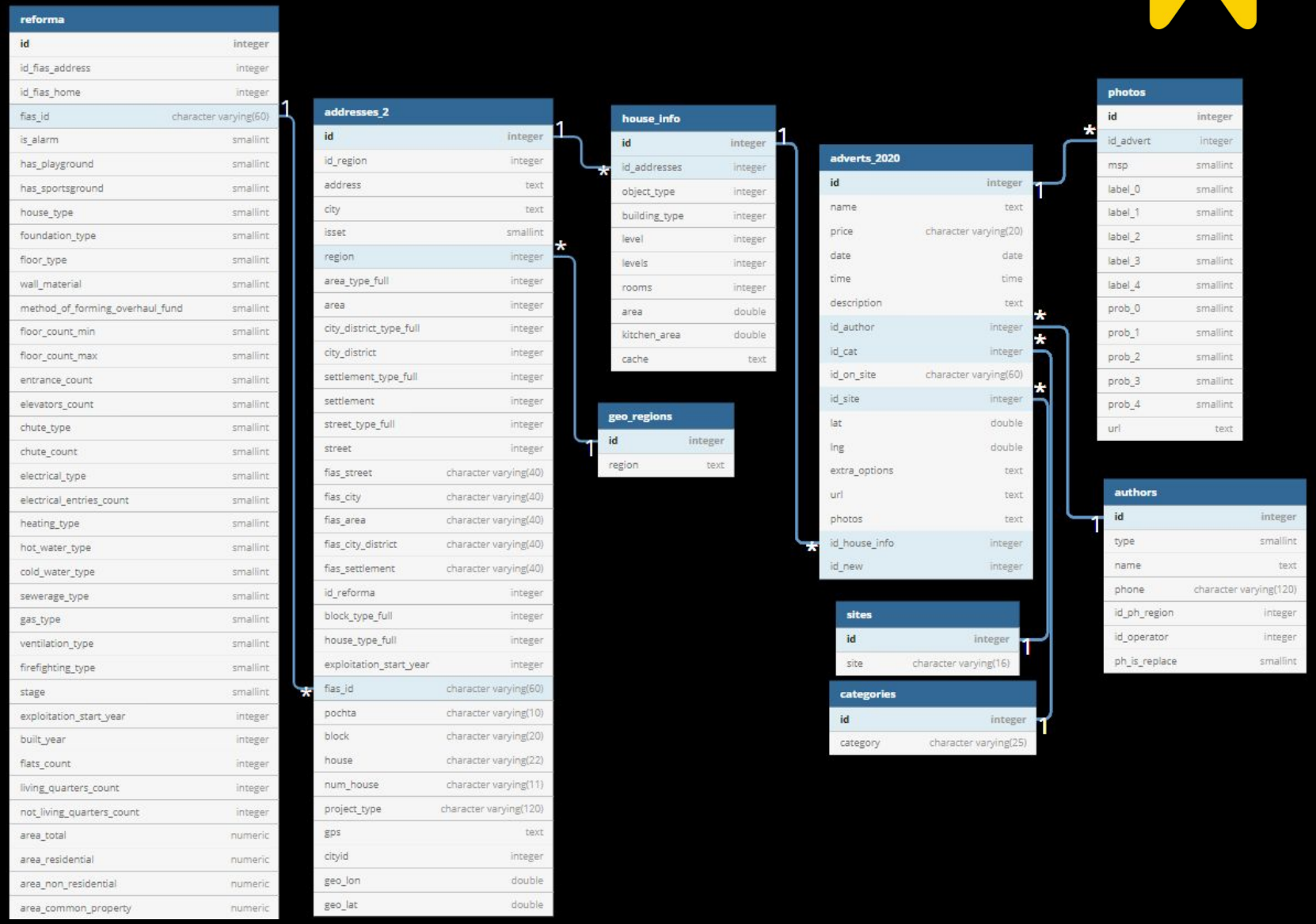
Размер таблиц

relation	size
adverts_2019	155 GB
adverts_2020	97 GB
adverts_2021	40 GB
adverts_2018	21 GB
re_adverts_20	21 GB
re_adverts_2021	15 GB
author_api	14 GB
houses	9655 MB
addresses_2	8751 MB
houses_2	5865 MB

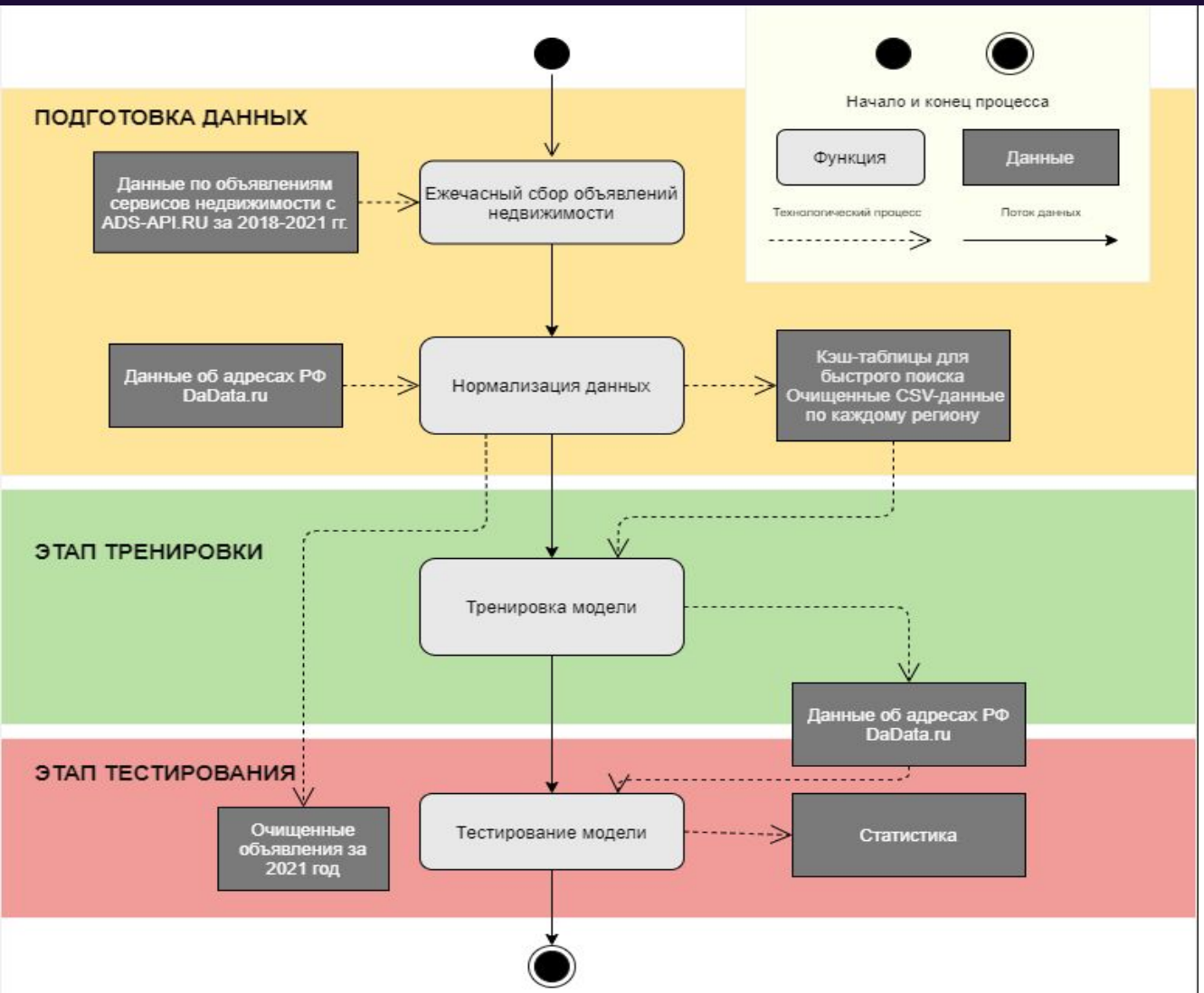


ДАТАСЕТ

Проработана архитектура базы данных на основе PostgreSQL



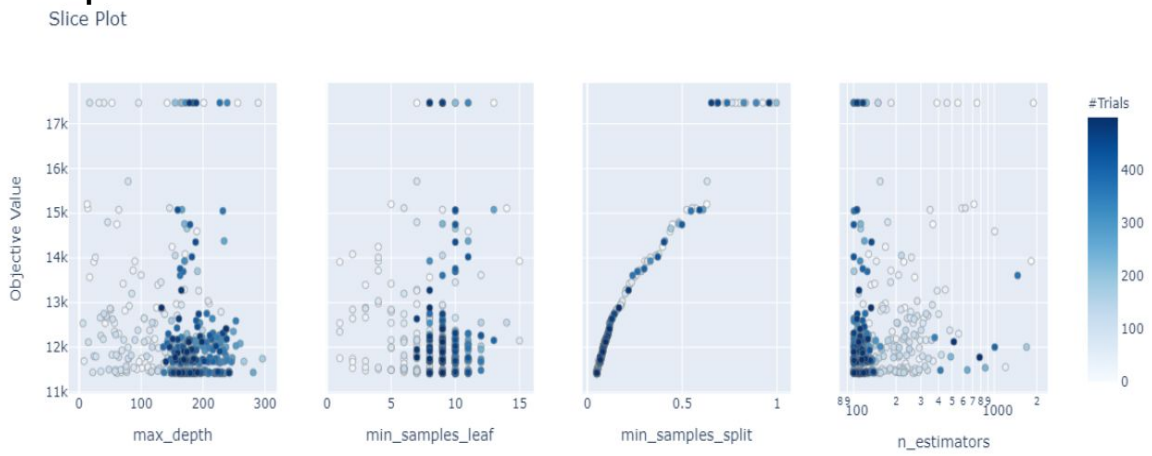
Архитектура и ГИПЕПАРАМЕТРЫ



Алгоритмы:

1. Random Forest (Случайный лес)
2. XGBoost
3. SQL-запросы: «Похожие» объекты и определение средних значений

Оптимизация гиперпараметров — задача машинного обучения по выбору набора оптимальных гиперпараметров для обучающего алгоритма



ЦЕНООБРАЗУЮЩИЕ ФАКТОРЫ

Используются на данный момент:

1. Площадь
2. Площадь кухни
3. Этаж
4. Этажность
5. Количество комнат
6. Тип дома: панельный, кирпичный...
7. Время публикации
8. Стоимость доллара на момент публикации
9. Расстояние от центра столицы региона (на основе координат)
10. Азимут от центра столицы региона
11. Цена
12. Рынок недвижимости: «Вторичка» или «Новостройка»
13. Регион
14. Данные с «Реформа ЖКХ»

Дополнительные параметры, которые используются экспериментально:

1. Средние показатели оценки по всем фотографиям объявления
2. Доступность в метрах к важным местам:
    1. Ближайшие остановки
    2. Ближайшее метро
    3. Инфраструктура рядом: школы, магазины

ИТОГОВАЯ СИСТЕМА

Оценка объекта недвижимости

В данной форме вы можете ввести характеристики недвижимости для её оценки. Для корректной оценки стоимости необходимо заполнить **все характеристики, отмеченные звёздочкой** .

Тип жилья *

Тип дома *

Количество этажей *

Введите количество этажей

Этаж *

5

Количество комнат *

Площадь квартиры (кв.м.) *

78,9

Площадь кухни (кв.м.) *

Введите площадь кухни

Адрес объекта: Республика Татарстан, г Набережные
49, кв 18

Изменить адрес

ОЦЕНИТЬ

История продаж

Дата публикации: все | 2021 | 2020 | 2019 | 2018

Количество комнат: все | 1 | 2 | 3 | 4

Квартира	Начальная цена	Конечная цена	
65 м², 2 комнатная 2 этаж	3 950 000 руб. 60 769 руб. за м²	3 950 000 руб. 60 769 руб. за м²	Опубликовано - 29.10.20 Дата снятия - 29.11.20
83.4 м², 3 комнатная	6 050 000 руб.	6 050 000 руб.	Опубликовано - 27.11.20

ИТОГОВАЯ СИСТЕМА

predict Нейронные сети

POST	/predict/byPhotos	Определение типа локации, уровеней эстетики и качества фотографии
POST	/predict/request	XGBoost, RandomForest, CatBoost
POST	/predict/requestByParam	XGBoost, RandomForest, CatBoost

trend Тенденция изменения

POST	/trend/getAverage	Похожие цены по параметрам объявлениям
POST	/trend/getPriceChange	Похожие цены по географической области

bedchamber	
skyscraper	
elevator/door	
kitchen	
roof_garden	
skyscraper	
bathroom	
television_room	

НАУЧНАЯ СТАТЬЯ

В статьей представлена работа с классификацией изображений при помощи модели ResNet34

Электронные библиотеки. 2020. Т. 23. № 6

УДК 004.8

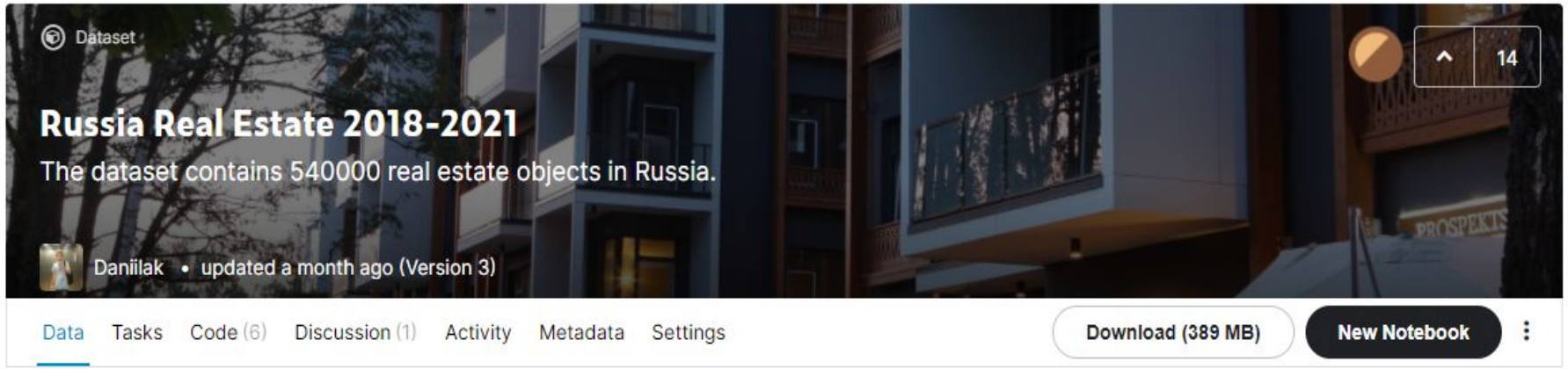
ОПРЕДЕЛЕНИЕ ДЕФЕКТОВ НА СТАЛЬНЫХ ЛИСТАХ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Р. Д. Гаскаров[1], А. М. Бирюков[2], А. Ф. Никонов[3], Д. В. Агниашвили[4], Д. А. Хайрисламов[5]

Институт информационных технологий и интеллектуальных систем, Казанский (Приволжский) федеральный университет, Казань

[1]komnatakita@gmail.com, [2]herou.public@gmail.com, [3]nikonowalex@gmail.com, [4]daniilak@gmail.com, [5]danilhayrislamov@gmail.com

ДАТАСЕТЫ



Usability 10.0 Licens

Usability 10.0 License GNU Affero General Public License 3.0 Tags real estate, housing, russia

ИТОГИ

author Работа с авторами объявлений

data Данные о справочниках

search Основные запросы, связанные с поиском по нормализированным DaData'ой объектам

POST `/search/getByGeo` Поиск объявлений по координатам

POST `/search/getByAddress` Поиск объявлений по адресу. Необходимо для поиска по домам

Возвращает массив идентификаторов объявлений

Parameters

Cancel

Name	Description
address * required string (query)	Адрес чебоксары, улица университетская 27
offset integer (query)	0
limit integer (query)	1000
radius integer (query)	Радиус поиска 0
object_type	Вид объекта: 1 - Вторичка; 2 - Новостройка; 3 - Машиноместо; 4 - Гараж; 5 - Торговое помещение; 6 - Помещение свободного назначения; 7 - Коммерческая

Спасибо за внимание!