

# 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.



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

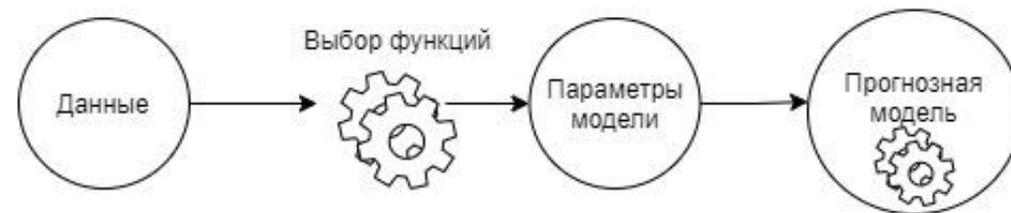
## Проблемы рынка недвижимости РФ:

- 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.

Статистическое /Эконометрическое моделирование:



Машинное обучение:



**Различия в методологиях моделирования**

## 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) **Oscenik.ru**

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

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

```
danilak@Debian-105-buster-64-minimal
OS: Debian GNU/Linux 10 (buster) x86_64
Host: 1.0
Kernel: 4.19.0-14-amd64
Uptime: 28 days, 11 hours, 3 mins
Packages: 1109 (dpkg)
Shell: zsh 5.7.1
Terminal: /dev/pts/6
CPU: AMD Ryzen 5 3600 6- (12) @ 3.600GHz
GPU: ASPEED Technology, Inc. ASPEED Graphics Family
Memory: 4954MiB / 64416MiB
```

Рабочий сервер

```
danilak@Server
OS: Gentoo/Linux x86_64
Kernel: 5.10.26-gentoo-dist
Uptime: 12 days, 1 hour, 9 mins
Packages: 854 (emerge)
Shell: zsh 5.8
Resolution: 1920x1080
Terminal: /dev/pts/1
CPU: AMD Ryzen 5 3400G (8) @ 3.700GHz
GPU: AMD ATI 04:00.0 Picasso
Memory: 17943MiB / 31981MiB
```

Домашний сервер



# Используемые API

## 1. ADS-API.ru

Парсер недвижимости России. Собирается с [avito.ru](https://www.avito.ru), [irr.ru](https://www.irr.ru), [realty.yandex.ru](https://realty.yandex.ru), [cian.ru](https://www.cian.ru), [sob.ru](https://www.sob.ru), [youla.io](https://youla.io), [n1.ru](https://n1.ru) и [moyareklama.ru](https://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

08

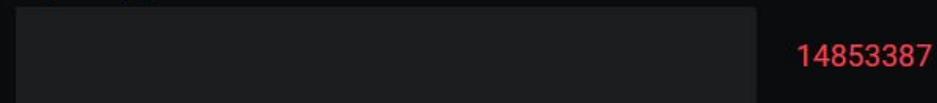
Размер таблиц	
relation	size
adverts_2019	155 GB
adverts_2020	97 GB
adverts_2021	40 GB
adverts_2018	21 GB
re_adverts_20	21 GB
re_adverts_2021	15 GB
author_api	14 GB
houses	9655 MB
addresses_2	8751 MB
houses_2	5865 MB

Общее количество объявлений

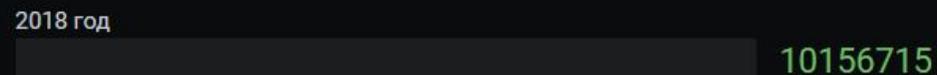
## Продажа



## Аренда



Количество объявлений по продаже за год



Количество объявлений по аренде за год



# ДАТАСЕТ

09

Проработана архитектура базы данных на основе PostgreSQL

reforma	
id	integer
id_fias_address	integer
id_fias_home	integer
fias_id	character varying(60)
is_alarm	smallint
has_playground	smallint
has_sportsground	smallint
house_type	smallint
foundation_type	smallint
floor_type	smallint
wall_material	smallint
method_of_forming_overhaul_fund	smallint
floor_count_min	smallint
floor_count_max	smallint
entrance_count	smallint
elevators_count	smallint
chute_type	smallint
chute_count	smallint
electrical_type	smallint
electrical_entries_count	smallint
heating_type	smallint
hot_water_type	smallint
cold_water_type	smallint
sewerage_type	smallint
gas_type	smallint
ventilation_type	smallint
firefighting_type	smallint
stage	smallint
exploitation_start_year	integer
built_year	integer
flats_count	integer
living_quarters_count	integer
not_living_quarters_count	integer
area_total	numeric
area_residential	numeric
area_non_residential	numeric
area_common_property	numeric

addresses_2	
id	integer
id_region	integer
address	text
city	text
isset	smallint
region	integer
area_type_full	integer
area	integer
city_district_type_full	integer
city_district	integer
settlement_type_full	integer
settlement	integer
street_type_full	integer
street	integer
fias_street	character varying(40)
fias_city	character varying(40)
fias_area	character varying(40)
fias_city_district	character varying(40)
fias_settlement	character varying(40)
id_reforma	integer
block_type_full	integer
house_type_full	integer
exploitation_start_year	integer
fias_id	character varying(60)
pochta	character varying(10)
block	character varying(20)
house	character varying(22)
num_house	character varying(11)
project_type	character varying(120)
gps	text
cityid	integer
geo_lon	double
geo_lat	double

house_info	
id	integer
id_addresses	integer
object_type	integer
building_type	integer
level	integer
levels	integer
rooms	integer
area	double
kitchen_area	double
cache	text

geo_regions	
id	integer
region	text

adverts_2020	
id	integer
name	text
price	character varying(20)
date	date
time	time
description	text
id_author	integer
id_cat	integer
id_on_site	character varying(60)
id_site	integer
lat	double
lng	double
extra_options	text
url	text
photos	text
id_house_info	integer
id_new	integer

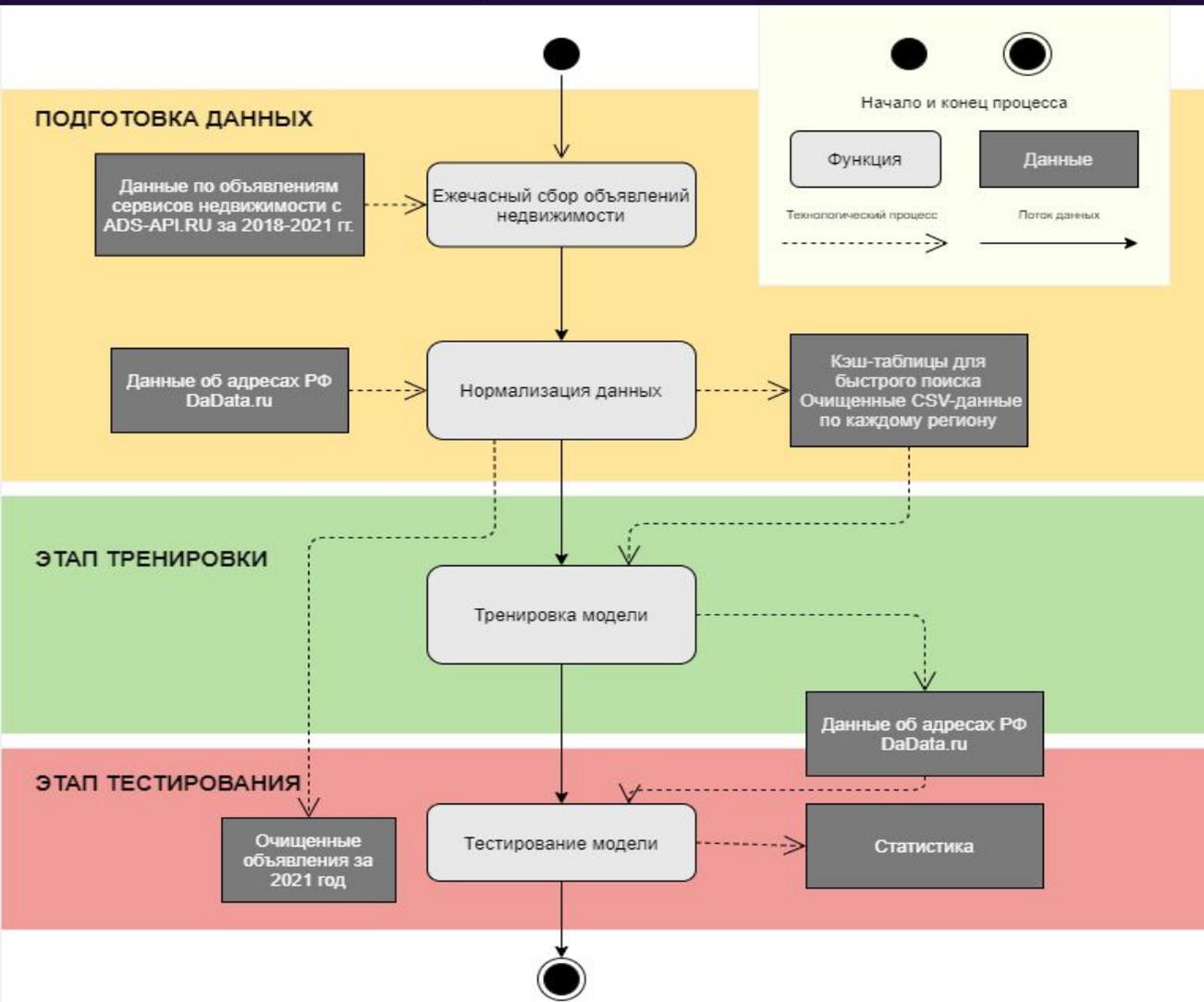
sites	
id	integer
site	character varying(16)

categories	
id	integer
category	character varying(25)

photos	
id	integer
id_advert	integer
misp	smallint
label_0	smallint
label_1	smallint
label_2	smallint
label_3	smallint
label_4	smallint
prob_0	smallint
prob_1	smallint
prob_2	smallint
prob_3	smallint
prob_4	smallint
url	text

authors	
id	integer
type	smallint
name	text
phone	character varying(120)
id_ph_region	integer
id_operator	integer
ph_is_replace	smallint



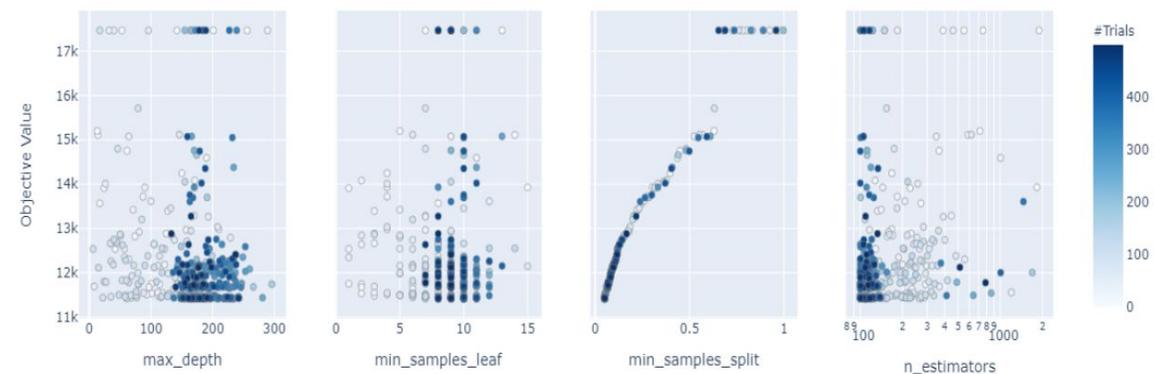


## Алгоритмы:

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

Slice Plot



## 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

## Оценка объекта недвижимости

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

Тип жилья \*  Тип дома \*

Количество этажей \*  Этаж \*

Количество комнат \*  Площадь квартиры (кв.м.) \*

Площадь кухни (кв.м.) \*

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

Изменить адрес

**ОЦЕНИТЬ**

## История продаж

Дата публикации

Количество комнат

	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



## predict Нейронные сети

**POST** /predict/byPhotos Определение типа локации, уровней 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	

В статье 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<sup>1</sup>, А. М. Бирюков<sup>2</sup>, А. Ф. Никонов<sup>3</sup>, Д. В. Агниашвили<sup>4</sup>,  
Д. А. Хайрисламов<sup>5</sup>

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

<sup>1</sup>komnatakita@gmail.com, <sup>2</sup>herou.public@gmail.com, <sup>3</sup>nikonowalex@gmail.com,  
<sup>4</sup>daniilak@gmail.com, <sup>5</sup>danilhayrislamov@gmail.com



Dataset

## Russia Real Estate 2018-2021

The dataset contains 540000 real estate objects in Russia.

Daniilak • updated a month ago (Version 3)

Data Tasks Code (6) Discussion (1) Activity Metadata Settings

Download (389 MB) New Notebook

Usability 10.0 License

Dataset

## Russia Real Estate & House Info 2018-2021

About 2,238,764 real estate objects in Russia with info about the object's house

Daniilak • updated 25 days ago

Data Tasks (1) Code Discussion Activity Metadata Settings

Download (1 GB) New Notebook

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



**author** Работа с авторами объявлений >

**data** Данные о справочниках >

**search** Основные запросы, связанные с поиском по нормализованным 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
<b>address</b> * required	Адрес
string (query)	<input type="text" value="чебоксары, улица университетская 27"/>
offset	0
integer (query)	
limit	1000
integer (query)	
radius	Радиус поиска.
integer (query)	0
object_type	Тип объекта: 1 - Регионы; 2 - Населенные пункты; 3 - Микрорайоны; 4 - Города; 5 - Типовые планировки; 6 - Планировки конкретных домов; 7 - Улицы/переулки



# Спасибо за внимание!



Казанский  
федеральный  
университет

ИНСТИТУТ  
информационных технологий  
и интеллектуальных систем

