

# СИСТЕМА АВТОМАТИЧНОГО ОЦІНЮВАННЯ НЕРУХОМОГО МАЙНА

**Науковий керівник**  
Дідковська М.В.

**Виконав**  
Іванов Даніїл

# МЕТА РОБОТИ

- 1) Сформувати достатню для роботи базу даних.
- 2) Проаналізувати існуючі методи побудови регресійних моделей.
- 3) Розробити програмний продукт для оцінки вартості квартир у місті Києві.

# АКТУАЛЬНІСТЬ



**Інвестиції**



**Час**



**Суб'єктивність**

# ПРИКЛАДИ ІСНУЮЧИХ РІШЕНЬ

1. АМЕРИКАНСЬКА КОМПАНІЯ ZILLOW

2. УКРАЇНСЬКА КОМПАНІЯ DOM.RIA

Розрахуйте вартість своєї квартири

В області \* Київська

В місті \* Київ

В районі Святошинський

Кількість кімнат \* 1 2 3 4 5+

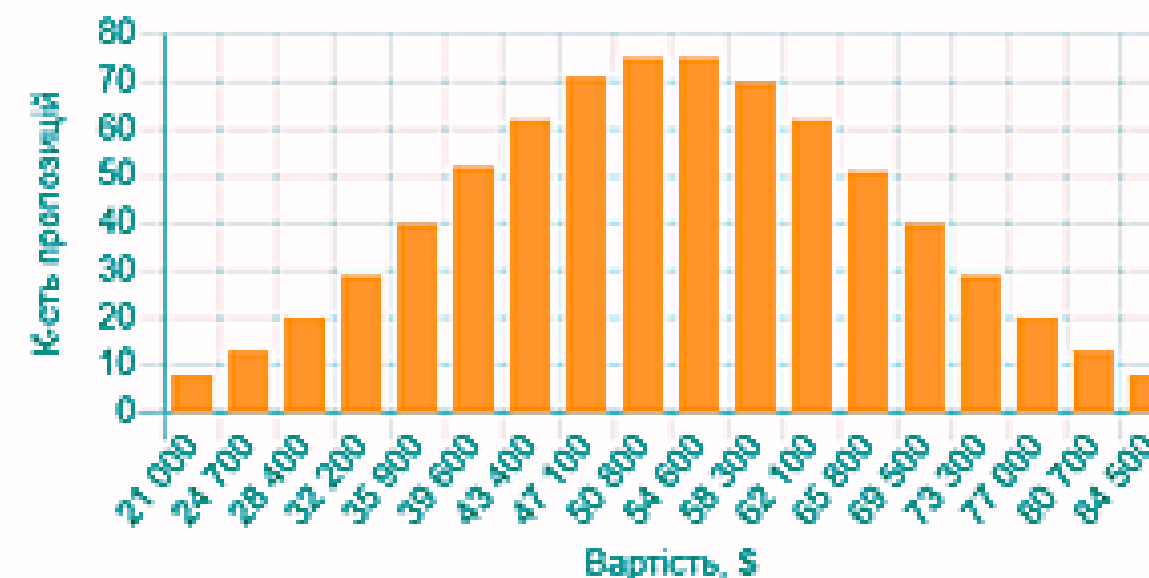
Загальна площа, м² \* 55

Розрахувати вартість

**49 490 \$**

Додати оголошення

(доларів США, станом на травень 2020)

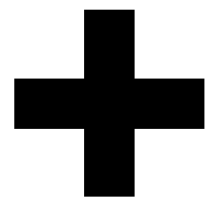


\*Виконано в якості консультаційної послуги на основі аналізу **757** подібних квартир

Онлайн-калькулятор від компанії DOM.RIA.

# ФОРМУВАННЯ БАЗИ ДАНИХ

Веб  
парсер



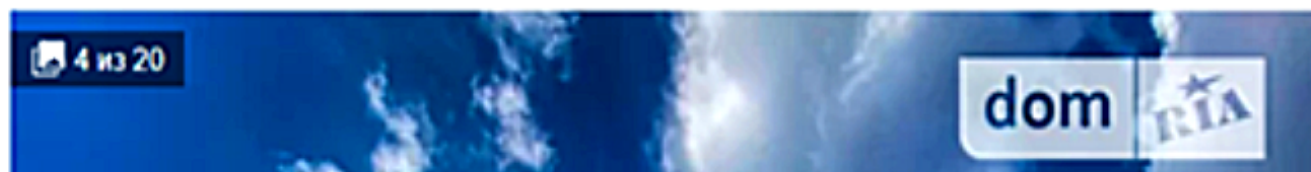
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 22791 entries, 0 to 22950
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   price_total_usd                       22791 non-null  float64
1   total_square_meters                   22791 non-null  float64
2   kitchen_square_meters                 16302 non-null  float64
3   living_square_meters                  14823 non-null  float64
4   heating                               22748 non-null  float64
5   building_year                         10355 non-null  float64
6   wall_type_uk                          22791 non-null  int64
7   floors_count                          22791 non-null  int64
8   floor                                 22791 non-null  int64
9   rooms_count                           22791 non-null  int64
10  publishing_date                       22791 non-null  object
11  latitude                              22791 non-null  float64
12  longitude                             22791 non-null  float64
13  realty_id                             22791 non-null  int64
14  url                                    22791 non-null  object
15  metro_distance                        22791 non-null  int64
16  center_distance                       22791 non-null  int64
17  price_metr                            22791 non-null  float64
18  azimuth                               22791 non-null  float64
dtypes: float64(10), int64(7), object(2)
memory usage: 3.5+ MB
```

# ОБРОБКА БАЗИ ДАНИХ

☆ Продаю 3к квартиру 116 кв. м, Драгомирова улица 14А  
возле метро Дружбы народов в районе Печерский в Киеве

**22 214 000 \$** за объект

600 378 444 грн • 191 500 \$ за м² [курс валют](#)



Торг Вторичное жилье

**Продажа квартиры в ЖК Покровский посад**

3 комнаты • 15 этаж из 23

Площадь 149000 м² • 56 м² • 46 м²

Торг Первичное жилье

**Продажа квартиры в ЖК Эврика**

3 комнаты • 23 этаж из 24

Площадь 76.53 м² • 45.47 м² • 892 м²

# МЕТОДИ ПОБУДОВИ РЕГРЕСІЙНИХ МОДЕЛЕЙ

## ЛІНІЙНА РЕГРЕСІЯ

### переваги:

- Добре вивчені
- Дуже швидкі
- Практично поза конкуренцією, коли ознак дуже багато

### недоліки:

- Багато вимог до даних
- Погано працюють зі складними нелінійними залежностями

# МЕТОДИ ПОБУДОВИ РЕГРЕСІЙНИХ МОДЕЛЕЙ

## ВИПАДКОВИЙ ЛІС

### переваги:

- Висока точність передбачення
- Майже не чутливий до викидів в даних
- Існують методи оцінювання значущості окремих ознак в моделі

### недоліки:

- Складніше інтерпретувати результати, на відміну від одного дерева
- Потребує багато пам'яті



# МЕТОДИ ПОБУДОВИ РЕГРЕСІЙНИХ МОДЕЛЕЙ

## ГРАДІЄНТНИЙ БУСТИНГ

### переваги:

- Може описати досить складну функцію
- В стандартних завданнях дуже часто є найефективнішим алгоритмом

### недоліки:

- Схильний до перенавчання

# АЛГОРИТМ

**Випадковий ліс**  
**+**  
**градієнтний бустинг**

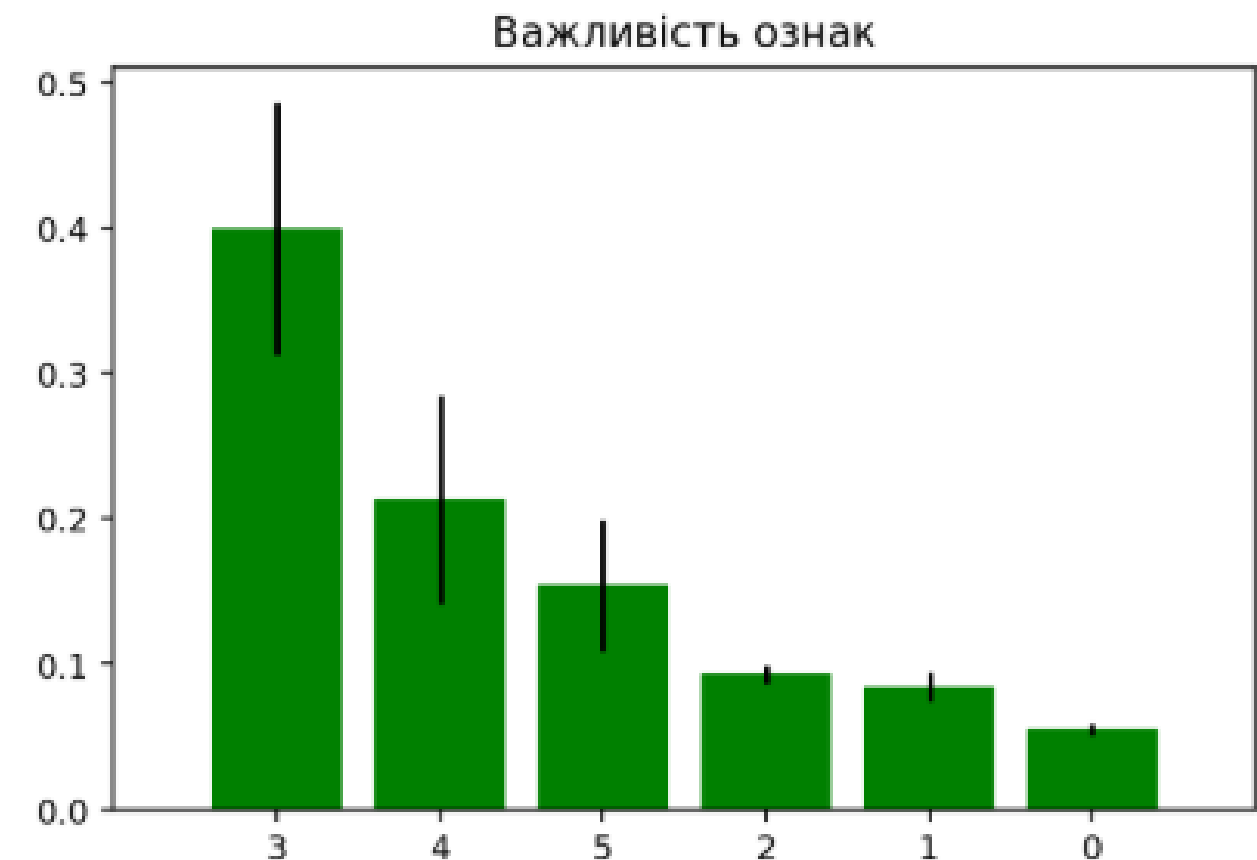
# АЛГОРИТМ

## Інтерпретація даних

Latitude + Longitude =

- center\_distance - відстань до центра міста
- metro\_distance - відстань до найближчої станції метрополітену
- azimuth - кут між напрямком на північ та на центр міста

```
1. center_distance (0.399699)
2. metro_distance (0.213236)
3. azimuth (0.154019)
4. total_square_meters (0.092887)
5. floors_count (0.084609)
6. floor (0.055550)
```



**Рейтинг важливості ознак в моделі Random forest.**

# ЗАСОБИ ДЛЯ РОЗРОБКИ

Мова програмування

**Python 3.8**

Збір бази даних

**requests – для взаємодії з веб-сайтом DOM.RIA**  
**re – для роботи з регулярними виразами**

Навчання моделей

**Jupyter Notebook**  
**Scikit-learn**

Кінцевий програмний  
продукт

**Django REST Framework**

# КРИТЕРІЇ ЯКОСТІ

## 1) Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\Phi_i - \Pi_i}{\Phi_i} \right|$$

## 2) MEDium Absolute Percentage Error (MedAPE)

## 3) Коефіцієнт детермінації $R^2$

# АНАЛІЗ ЯКОСТІ РОБОТИ СИСТЕМИ

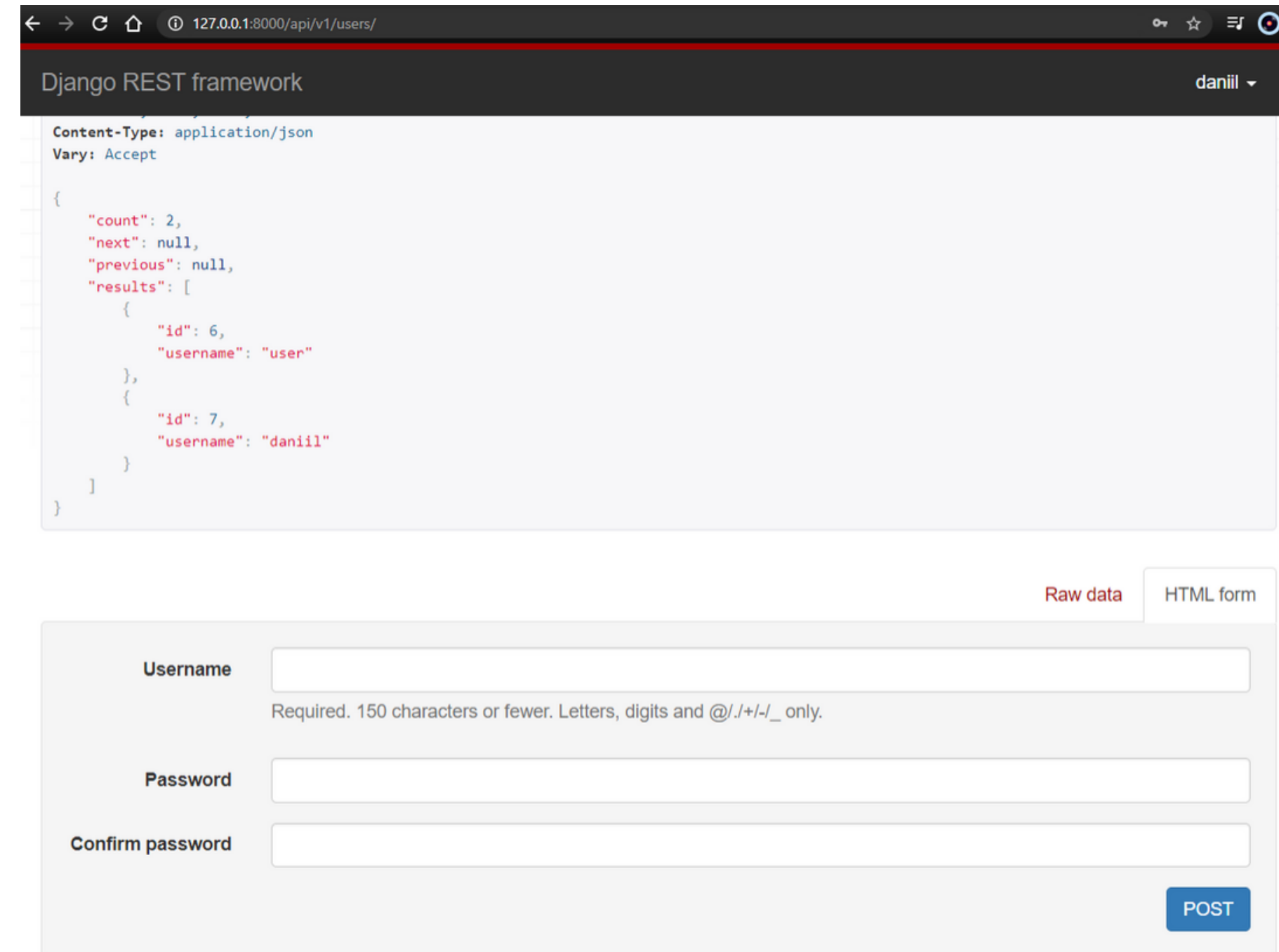
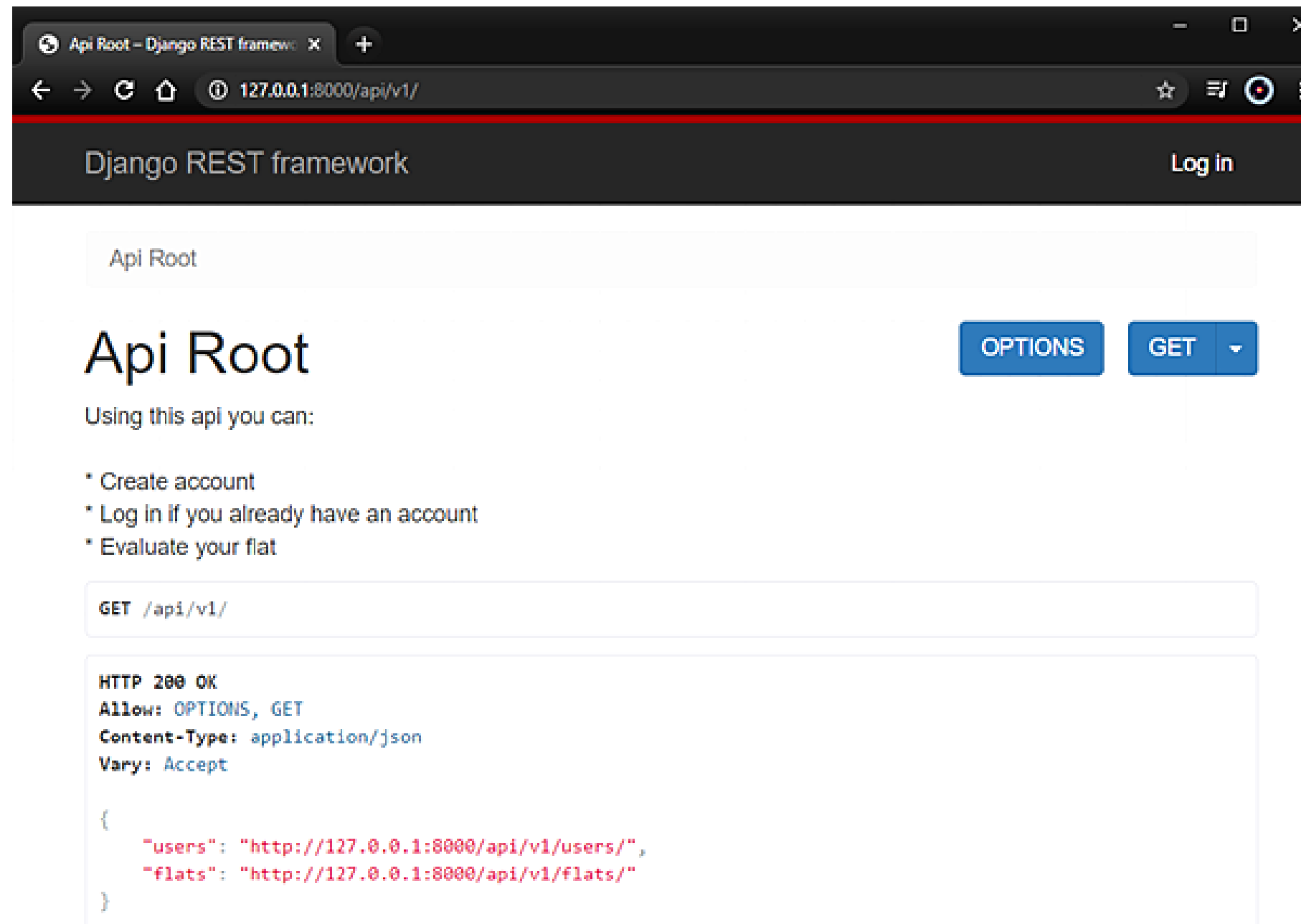
$$\text{MAPE} = 12.2\%$$

$$\text{MedAPE} = 4.2\%$$

$$R^2 = 0.84$$

# КЕРІВНИЦТВО КОРИСТУВАЧА API

## БРАУЗЕРНА ВЕРСІЯ



# КЕРІВНИЦТВО КОРИСТУВАЧА API

## БРАУЗЕРНА ВЕРСІЯ

Django REST framework

danil

Api Root / Flat List Create Api

### Flat List Create Api

OPTIONSGET

This view allows to Retrieve all your flats or to add a new one.

Data requirements:

- \* flat should be located not farther then 20 km from Kyiv city center;
- \* "floors count" >= 4 and <= 48
- \* "floor" >= 1 and <= 48
- \* "total square meters" >= 20 and <= 200

GET /api/v1/flats/

HTTP 200 OK

Allow: GET, POST, HEAD, OPTIONS

Content-Type: application/json

Vary: Accept

```
{
  "count": 3,
  "next": null,
  "previous": null,
  "results": [
    {
      "edit_flat": "http://127.0.0.1:8000/api/v1/flats/12/",
      "user": "user",
      "floor": 5,
      "floors_count": 7,
      "total_square_meters": 183.0,
      "latitude": 50.424018,
      "longitude": 30.520902,
      "center_distance": 2907,
      "metro_distance": 293,
    }
  ]
}
```

Django REST framework

danil

```
{
  "floor": 4,
  "floors_count": 9,
  "total_square_meters": 183.0,
  "latitude": 50.424018,
  "longitude": 30.520902,
  "center_distance": 2907,
  "metro_distance": 293,
  "predicted_price_metr": 1809,
  "predicted_price_total": 331053
}
```

Raw data

HTML form

Floor

Floors count

Total square meters

Latitude

Longitude

POST



# КЕРІВНИЦТВО КОРИСТУВАЧА API

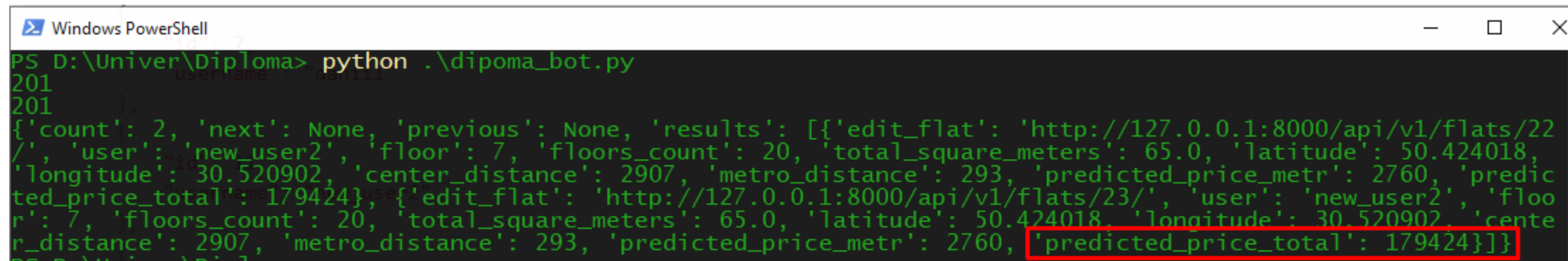
## ПРОГРАМНА ВЗАЄМОДІЯ

```
import requests
import json

url = 'http://127.0.0.1:8000/api/v1/users/'
response = requests.post(url, data={
    'username': 'new_user',
    'password': 'mypassword12',
    'confirm_password': 'mypassword12'})
print(response.status_code)

url = 'http://127.0.0.1:8000/api/v1/flats/'
response = requests.post(
    url,
    data={'floor': 7, 'floors_count': 20, 'total_square_meters': 65,
        'latitude': 50.424018, 'longitude': 30.520902},
    auth=('new_user2', 'mypassword12')
)
print(response.status_code)

response = requests.get(url, auth=('new_user2', 'mypassword12'))
print(response.json())
```



```
Windows PowerShell
PS D:\Univer\Diploma> python .\dipoma_bot.py
201
201
{'count': 2, 'next': None, 'previous': None, 'results': [{'edit_flat': 'http://127.0.0.1:8000/api/v1/flats/22/', 'user': 'new_user2', 'floor': 7, 'floors_count': 20, 'total_square_meters': 65.0, 'latitude': 50.424018, 'longitude': 30.520902, 'center_distance': 2907, 'metro_distance': 293, 'predicted_price_metr': 2760, 'predicted_price_total': 179424}, {'edit_flat': 'http://127.0.0.1:8000/api/v1/flats/23/', 'user': 'new_user2', 'floor': 7, 'floors_count': 20, 'total_square_meters': 65.0, 'latitude': 50.424018, 'longitude': 30.520902, 'center_distance': 2907, 'metro_distance': 293, 'predicted_price_metr': 2760, 'predicted_price_total': 179424}]]
```

# КЕРІВНИЦТВО КОРИСТУВАЧА API

## ПРОГРАМНА ВЗАЄМОДІЯ

### Freemium

Безкоштовна версія:  
**>= 500 запитів на день**

Платна версія:  
**без обмежень**

```
GET /api/v1/
```

```
HTTP 429 Too Many Requests
```

```
Allow: GET, OPTIONS
```

```
Content-Type: application/json
```

```
Retry-After: 86388
```

```
Vary: Accept
```

```
{  
  "detail": "Request was throttled. Expected available in 86388 seconds."  
}
```

*Приклад помилки №429.*

# ВИСНОВКИ

- Проведено аналіз актуальності задачі оцінки вартості нерухомого майна, досліджено існуючі рішення.
- Проведено аналіз існуючих методів машинного навчання для розв'язку задачі регресії.
- Зібрано власну базу даних.
- Розроблено алгоритм та побудовано за ним модель оцінки вартості квартир з медіанною абсолютною процентною помилкою, рівною 4,2%.
- Створено прикладний програмний інтерфейс.

# ШЛЯХИ ПОДАЛЬШОГО РОЗВИТКУ

- Розміщення API на сервері для відкритого користування.
- Використання даних про вже продані об'єкти.
- Розробка окремої системи для оцінки комерційної нерухомості.
- Розробка версій для використання в інших містах України.

ДЯКУЮ ЗА УВАГУ!