

**2026**

**ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ БУДІВЕЛЬ  
ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ**

**ЗАХИСТ**

**РОБОТА СТУДЕНТА КН-4  
ТЕЛЄГІН Н. А.**



# ЗМІСТ

Дана **РОБОТИ** дипломна робота є розширенням попередньої курсової роботи

## **В межах курсової роботи було:**

1. Досліджено особливості прогнозування енергоспоживання.
2. Оглянуто теоретичні засади прогнозування енергоспоживання як частина задачі прогнозування часових рядів.
3. Обрано низку моделей різного типу.
4. Проведено порівняльний аналіз в межах задачі прогнозування енергоспоживання будівель конкретного міста Дейтон, США.
5. Створено власну гібридну модель, що показала кращі результати, аніж будь-яка з попередніх моделей.

## **В межах дипломної роботи було:**

1. Визначено попередні результати.
2. Оглянуто фактори впливу на прогнозування енергоспоживання та методологію у розв'язанні задачі.
3. Покращено моделі, додаванням підбору гіперпараметрів, покращено архітектуру моделей.
4. Інтегровано новий набір даних, що побудований на відкритому метеорологічному API
5. Створено програму, що здатна прогнозувати енергоспоживання для будь-якого міста за коректних вхідних умов.

# ПОПЕРЕДНІ РЕЗУЛЬТАТИ

|                    | Linear Regression | Random Forest  | XGBoost        | LSTM           |
|--------------------|-------------------|----------------|----------------|----------------|
| <b>R2</b>          | 0,7012            | 0,8611         | 0,8846         | 0,8846         |
| <b>Summer RMSE</b> | 1422.34 (11.29%)  | 983.60 (7.81%) | 621.46 (4.93%) | 940.12 (7.64%) |
|                    |                   |                |                |                |

Найкраща модель - **XGBoost**,  
найслабша - звичайна **лінійна**  
**регресія**.

На основі отриманих результатів було вирішено  
побудувати **гібридну модель**, що поєднає сильні  
сторони XGBoost та LSTM моделей.

## ПРОБЛЕМАТИКА ПОСТАВЛЕНОЇ ЗАДАЧІ:



## ОГЛЯНУТІ РІШЕННЯ:

1. Використання ерофотознімків та зображень з рівня вулиці з використанням CNN
2. Класичні ML-алгоритми (Random Forest, SVM, Gradient Boosting)
3. Гібридні моделі CNN–LSTM

## РЕАЛІЗОВАНІ РІШЕННЯ:

XGBoost, Gibrid model (stacking-ансамбль xgb\_main,, extra\_trees і ridge)

### LSTM МОДЕЛЬ

RMSE = 702.81 MW  
MAE = 497.08 MW  
 $R^2 = 0.8768$



### XGBOOST МОДЕЛЬ

RMSE = 647.28 MW  
MAE = 443.16 MW  
 $R^2 = 0.8955$



**GIBRID MODEL: RMSE = 615.36 MW,  
MAE = 424.82 MW,  $R^2 = 0.9055$ .**

Гібридна модель покращила всі метрики, зокрема зниження RMSE на 5% порівняно з XGBoost та на 12.4% порівняно з LSTM до покращень.

**БУЛО РЕАЛІЗОВАНО:**

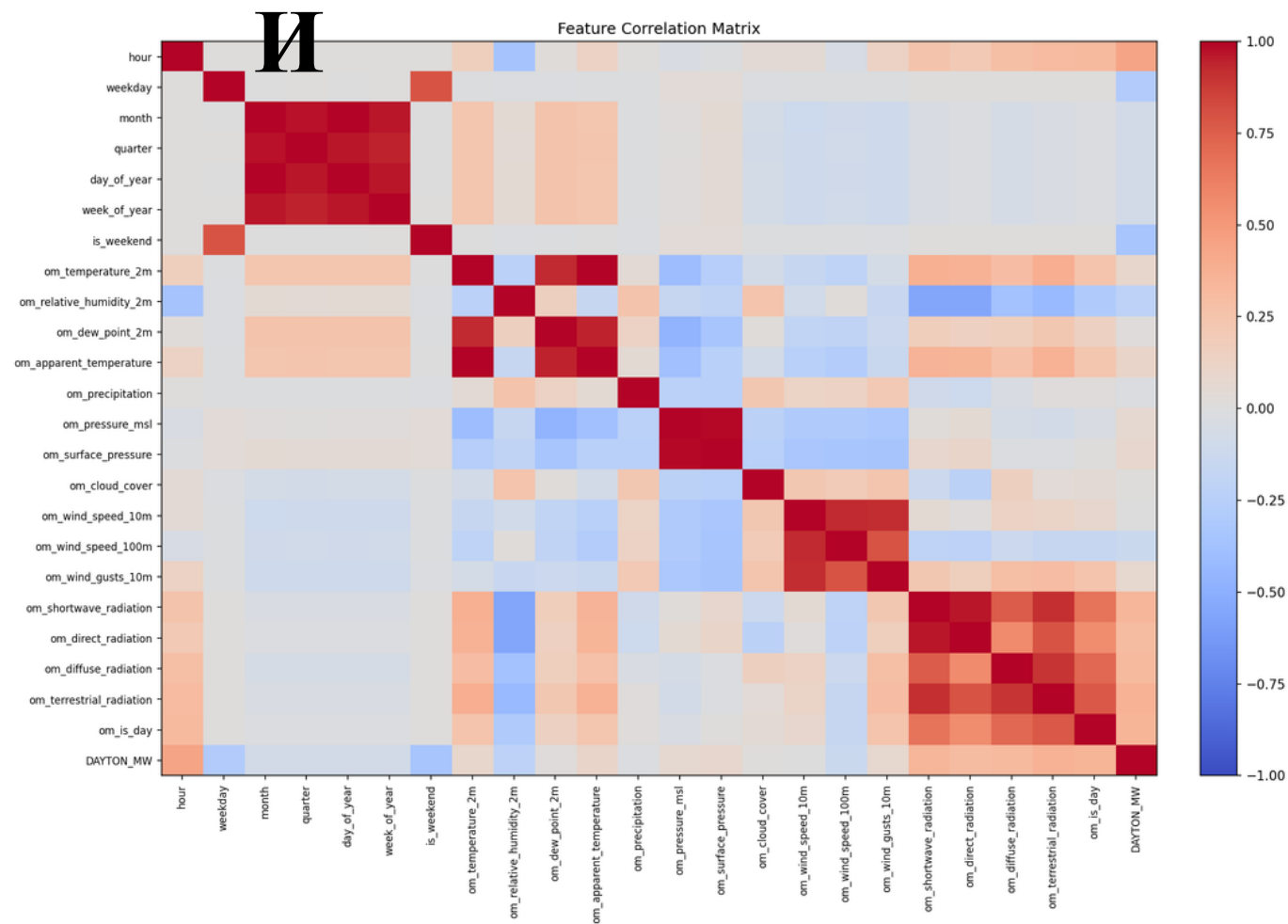
- 1. Інтеграція з відкритим метео API
- 2. Апроксимація та агрегування ознак.



| Datetime         | DAYTON_MW | Temp (°C) | Humidity (%) | Apparent Temp | Wind 10m | Pressure | Cloud Cover | Is Day | Radiation | Location |
|------------------|-----------|-----------|--------------|---------------|----------|----------|-------------|--------|-----------|----------|
| 2004-12-31 01:00 | 1596      | 10,8      | 77           | 6,2           | 7,12     | 1020,7   | 100         | 0      | 0         | Dayton   |
| 2004-12-31 02:00 | 1517      | 11        | 76           | 6,4           | 7,05     | 1020,3   | 100         | 0      | 0         | Dayton   |
| 2004-12-31 03:00 | 1486      | 11        | 76           | 6,5           | 7,12     | 1020,3   | 97          | 0      | 0         | Dayton   |
| 2004-12-31 04:00 | 1469      | 10,8      | 77           | 6,2           | 7,19     | 1019,9   | 100         | 0      | 0         | Dayton   |
| 2004-12-31 05:00 | 1472      | 10,9      | 79           | 6,5           | 7,09     | 1019,9   | 99          | 0      | 0         | Dayton   |

Усього маємо 50+ фіч, отриманих та спроектованих в межах датасету на основі даних з <https://open-meteo.com/>. Тобто, по-перше, датасет зріс майже удвічі, по-друге не є більше частковим випадком, а може бути сформований для будь-якого міста/локації.

# РЕЗУЛЬТАТ



## НАЙСИЛЬНІШІ ПОЗИТИВНІ ЗВ'ЯЗКИ:

hour → 0.44

om\_terrestrial\_radiation → 0.36

om\_is\_day → 0.35

om\_shortwave\_radiation → 0.34

## НАЙСИЛЬНІШІ НЕГАТИВНІ ЗВ'ЯЗКИ:

IS\_weekend → -0.34

weekday → -0.28

humidity → -0.21

Аналіз кореляційної матриці показав, що енергоспоживання найбільше залежить від часових факторів та рівня сонячної радіації. Водночас, значний вплив мають поведінкові фактори, такі як вихідні дні, що знижують рівень споживання. Погодні параметри, зокрема температура, виявились менш значущими, ніж очікувалось.

# РЕЗУЛЬТАТ И

| Feature                 | Imp |
|-------------------------|-----|
| is_weekend              |     |
| om_apparent_temperature |     |
| om_temperature_2m       |     |
| hour                    |     |
| weekday                 |     |

| Feature                  | Importance |
|--------------------------|------------|
| om_terrestrial_radiation | 0,0446     |
| om_is_day                | 0,0208     |
| month                    | 0,019      |
| week_of_year             | 0,0148     |
| day_of_year              | 0,0129     |

| Variant       | Validation RMSE | Test RMSE | MAE    | R²     |
|---------------|-----------------|-----------|--------|--------|
| Previous Best | —               | 615,36    | 424,82 | 0,9055 |
| New Base      | 79,39           | 82,96     | 61,77  | 0,9293 |
| New Full      | 18,53           | 18,94     | 13,92  | 0,9963 |

Під базовою моделлю вважається модель навчена без часових ознак, лише на метеорологічних ознаках, повна версія - з feature engeneering

Було побудовано програму, що для будь-якого міста могла б давати прогноз енергоспоживання на тиждень при наявності датасету, що складається з кортежу {HOUR, CONSUMING}

# РЕЗУЛЬТАТ



Було спроектовано застосунок, що дає змогу спрогнозувати енергоспоживання на основі:

- а) визначених конкретних ознак
- б) підвантаженого періоду прогнозу

Фактично, від обмеженої моделі з власноруч спроектованим датасетом і моделлю з точністю 90,55%, ми перейшли до моделі, здатна працювати з будь-яким містом (через API) і з вищою точністю при наявності даних про історичне енергоспоживання даного міста

Погодинні значення прогнозу

↓

🔍

⌵

|    | Datetime         | XGBoost  | Hybrid   |
|----|------------------|----------|----------|
| 0  | 2018-01-01 00:00 | 2,392.13 | 2,398.17 |
| 1  | 2018-01-01 01:00 | 2,330.81 | 2,354.36 |
| 2  | 2018-01-01 02:00 | 2,300.00 | 2,348.31 |
| 3  | 2018-01-01 03:00 | 2,297.63 | 2,350.32 |
| 4  | 2018-01-01 04:00 | 2,310.10 | 2,357.37 |
| 5  | 2018-01-01 05:00 | 2,353.04 | 2,390.29 |
| 6  | 2018-01-01 06:00 | 2,450.08 | 2,487.24 |
| 7  | 2018-01-01 07:00 | 2,638.01 | 2,614.38 |
| 8  | 2018-01-01 08:00 | 2,709.62 | 2,726.98 |
| 9  | 2018-01-01 09:00 | 2,721.74 | 2,756.36 |
| 10 | 2018-01-01 10:00 | 2,754.32 | 2,805.25 |
| 11 | 2018-01-01 11:00 | 2,811.84 | 2,823.16 |
| 12 | 2018-01-01 12:00 | 2,819.10 | 2,821.48 |

Оберіть тиждень 2018 року

2018-01-01 -> 2018-01-07 | ISO week 1

XGBoost

2,431.50 MW

↑ max 2,922.03

Hybrid

2,427.16 MW

↑ max 2,905.16

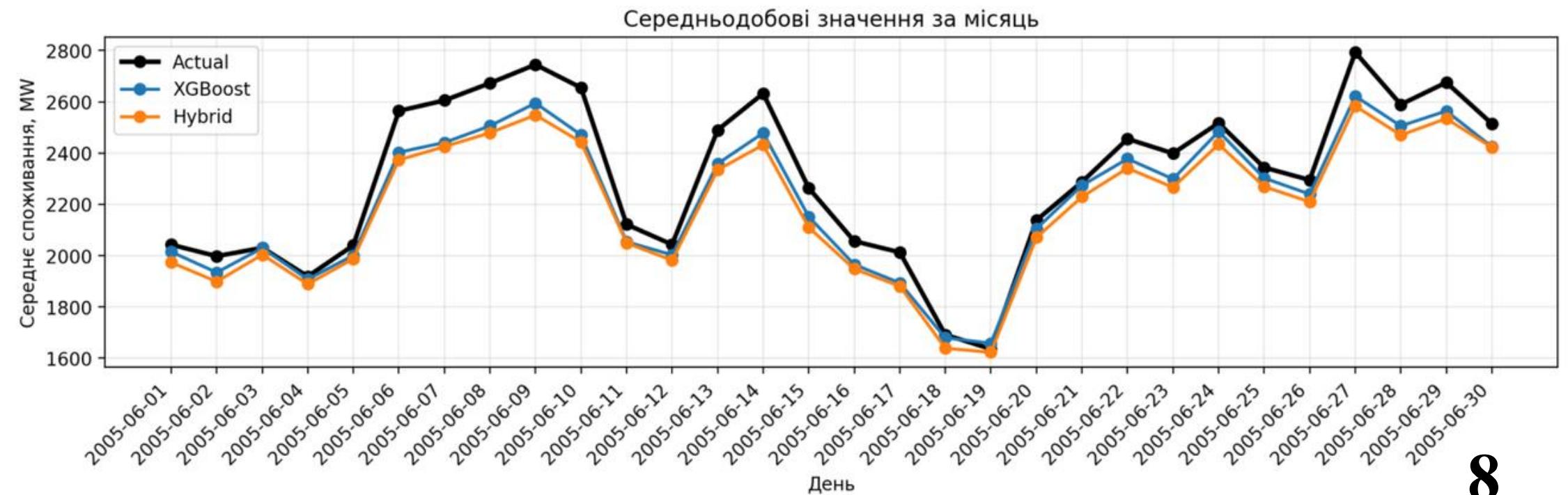
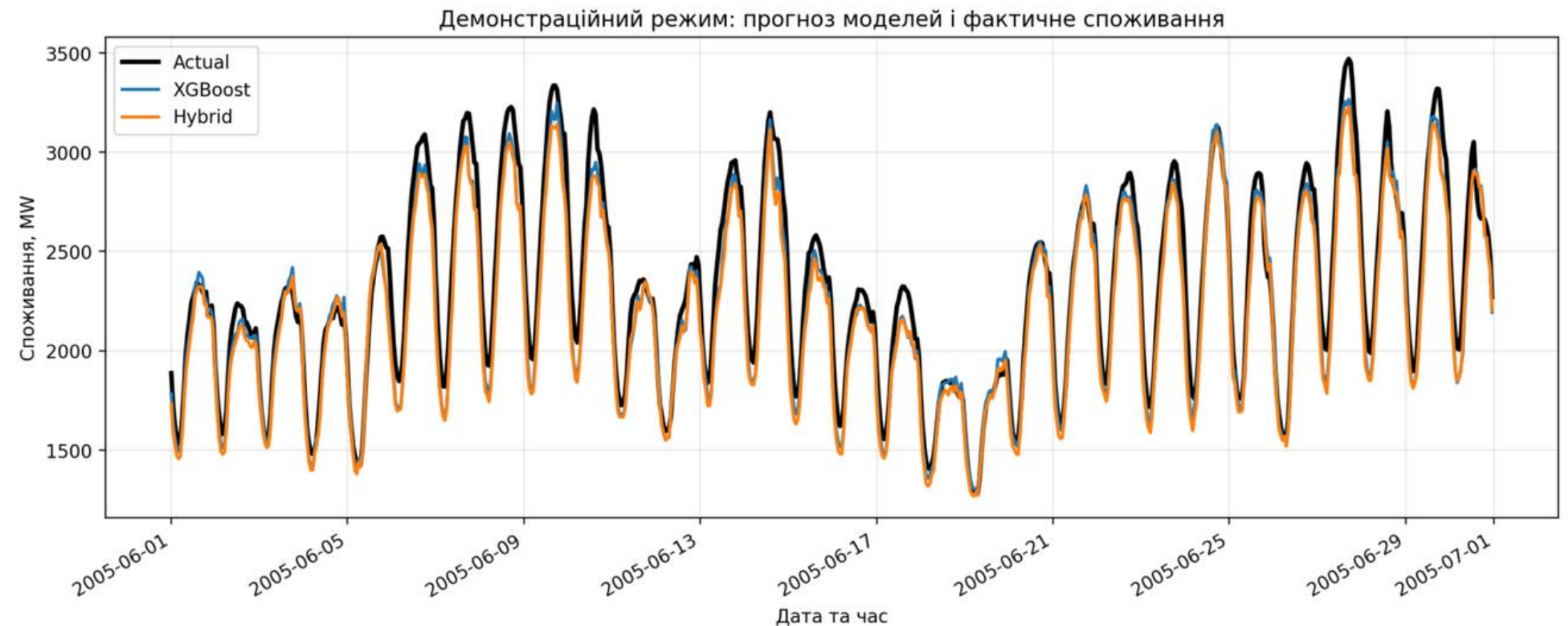
Порівняння моделей

|   | Модель  | Середній прогноз за тиждень, MW |
|---|---------|---------------------------------|
| 0 | XGBoost | 2,431.50                        |
| 1 | Hybrid  | 2,427.16                        |

# РЕЗУЛЬТАТИ

Користувач отримує:  
Подвійний прогноз: моделі  
XGBoost та Gybrid  
Уявлення про пікові  
навантаження та конкретні  
цифри.

**ЯК МОЖЕ БУТИ  
ВИКОРИСТАНО:**  
Створення оптимальних  
розподільчих мереж



## ДО ЗАХИСТУ ПЛАНУЄТЬСЯ:

1. Покращити структуру та оформлення роботи
2. Покращити структуру проєкту на GitHub
3. Спробувати використати альтернативні моделі машинного навчання у поставленій задачі.
4. Оглянути проблематику реалізації на рівні вулиць/мікрорайонів
5. Для об'єктивності результатів протестувати в межах декількох міст

## ПОСИЛАННЯ НА GИT ПРОЄКТУ:

[https://github.com/Nikita-ukma/Buildings\\_consumption](https://github.com/Nikita-ukma/Buildings_consumption)

**ДЯКУЮ ЗА УВАГУ**

