

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ ЕЛЕКТРООБЛАДНАННЯ НА ОСНОВІ ІНТЕРНЕТУ РЕЧЕЙ

Виконав студент групи 2ІСТ-19м
Науменко Денис Сергійович
науковий керівник: д.т.н., проф. Мокін О. Б.

- **Об'єктом** дослідження є процес споживання електричної енергії електрообладнанням будівлі з низьким енергоспоживанням.
- **Предметом** є інформаційна технологія прогнозування енергоспоживання електрообладнання будівлі з низьким енергоспоживанням.
- **Метою** є покращення енергоефективності будівлі з низьким енергоспоживанням за рахунок прогнозування потреб у енергії електрообладнання будівлі та вибору оптимального режиму роботи цього електрообладнання з використанням даних інтернету речей.

Основні задачі дослідження:

- прогнозувати показники енергоспоживання та ефективно розподіляти роботу електрообладнання будівлі з низьким енергоспоживанням;
- підбирати кращі тарифні плани серед тих, що пропонують електроенергетичні компанії;
- оцінити рівень енергоефективності будівлі з урахуванням показників зовнішнього середовища, отриманого за допомогою інтернету речей;

Актуальність

У сучасному світі, що швидко зростає та розвивається, проблема ефективного використання енергії стає все більш актуальною. Оптимізація витрат є одним з ключових викликів сучасного світу. Одним з напрямків такої оптимізації є створення інформаційних технологій прогнозування витрат енергії, які дозволяють краще розуміти потреби у енергії та ефективніше підбирати джерела її генерації. Це і зумовлює актуальність обраного напрямку дослідження, а впровадження та розвиток смарт-технологій та інтернету речей дозволяють перевести інформаційні технології такого типу на новий рівень аналізу та прогнозування.

Існуючі Аналоги

- Інформаційна технологія з використанням LSTM, ARIMA та бібліотеки Prophet
- Інформаційна технологія з використанням LSTM, та рекурентної нейронної мережі
- Інформаційна технологія з використанням k-найближчих сусідів

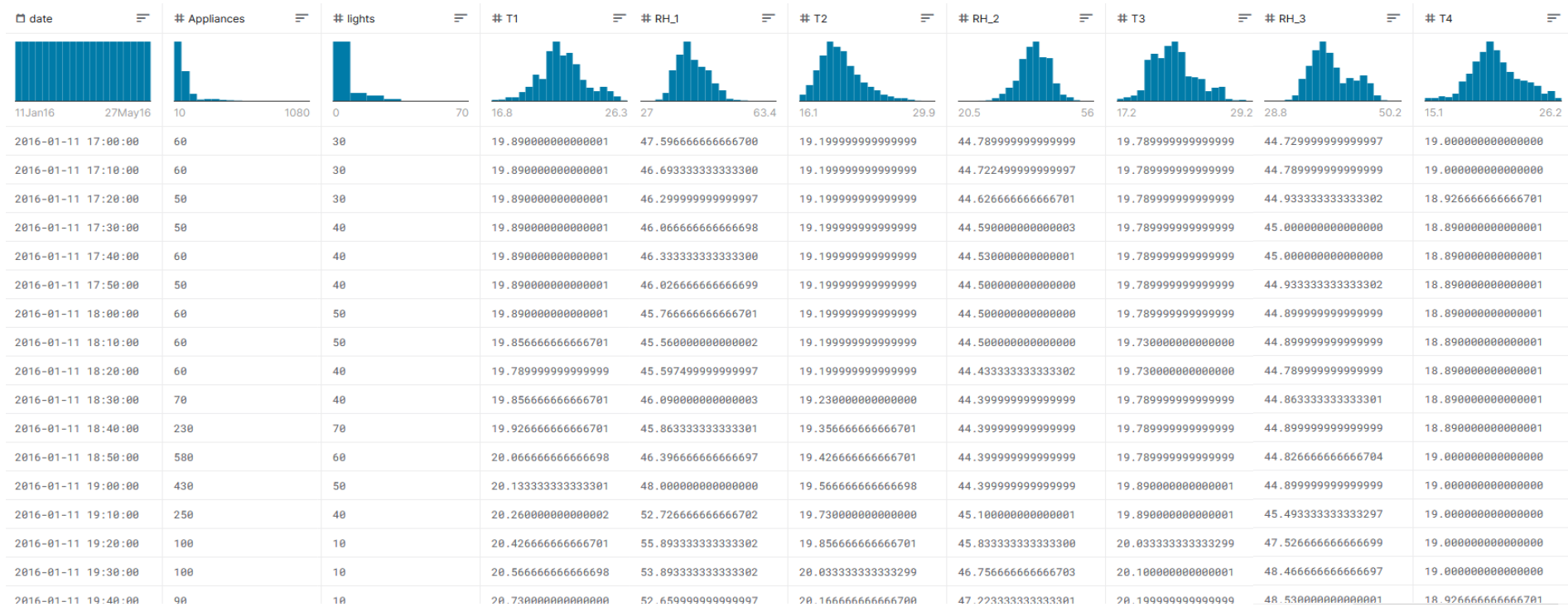
Елементи технології

- Scikit-learn (також відома як sklearn або scikits.learn) — це безкоштовна програмна бібліотека машинного навчання для мови програмування Python, яка надає функціональність для створення та тренування різноманітних алгоритмів класифікації, регресії та кластеризації, таких як лінійна регресія, random forest, градієнтний бустинг, і працює у зв'язці з бібліотеками NumPy та SciPy.
- Seaborn - це бібліотека для створення статистичної графіки на Python.
- NumPy - це основний пакет, необхідний для наукових обчислень з Python.

Kaggle

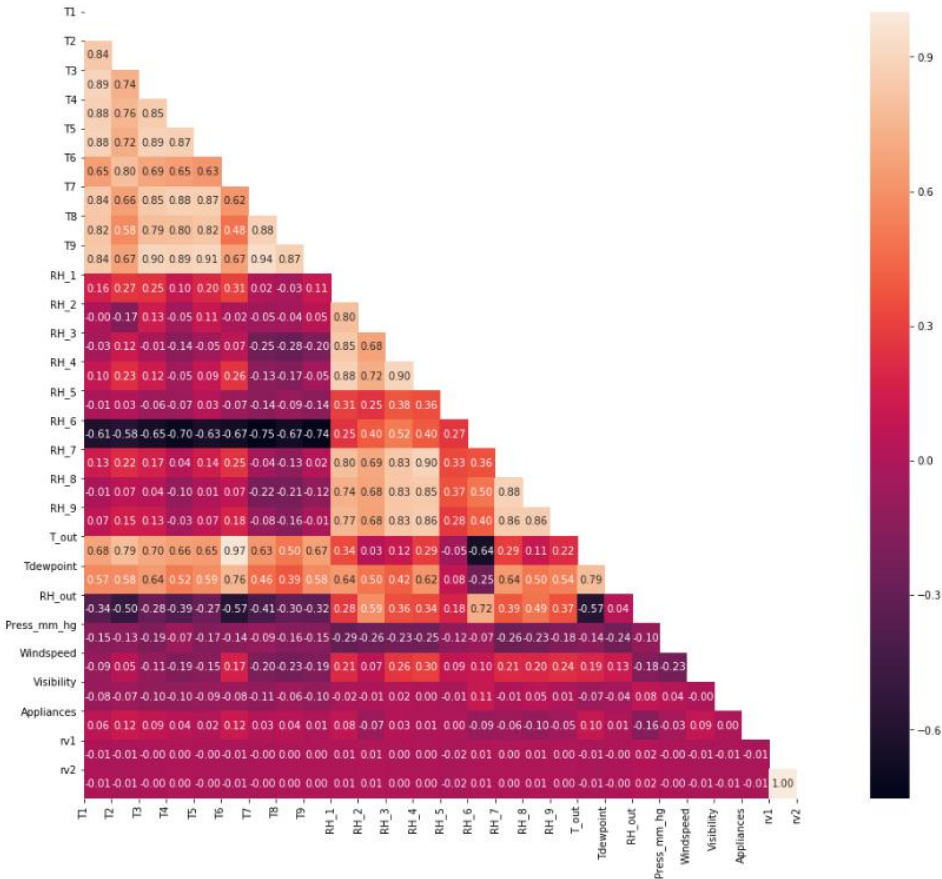
Це платформа для змагань з аналітики та передбачувального моделювання, в рамках якого статистики та добувачі даних конкурують у створенні найкращих моделей для прогнозування та опису даних, запропонованих компаніями або користувачами. Цей краудсорсинговий підхід ґрунтується на тому, що є безліч стратегій, які можуть бути застосовані до будь-якого завдання з передбачувального моделювання, і наперед не відомо, яка методика або аналітичний підхід буде найбільш ефективним

Огляд даних



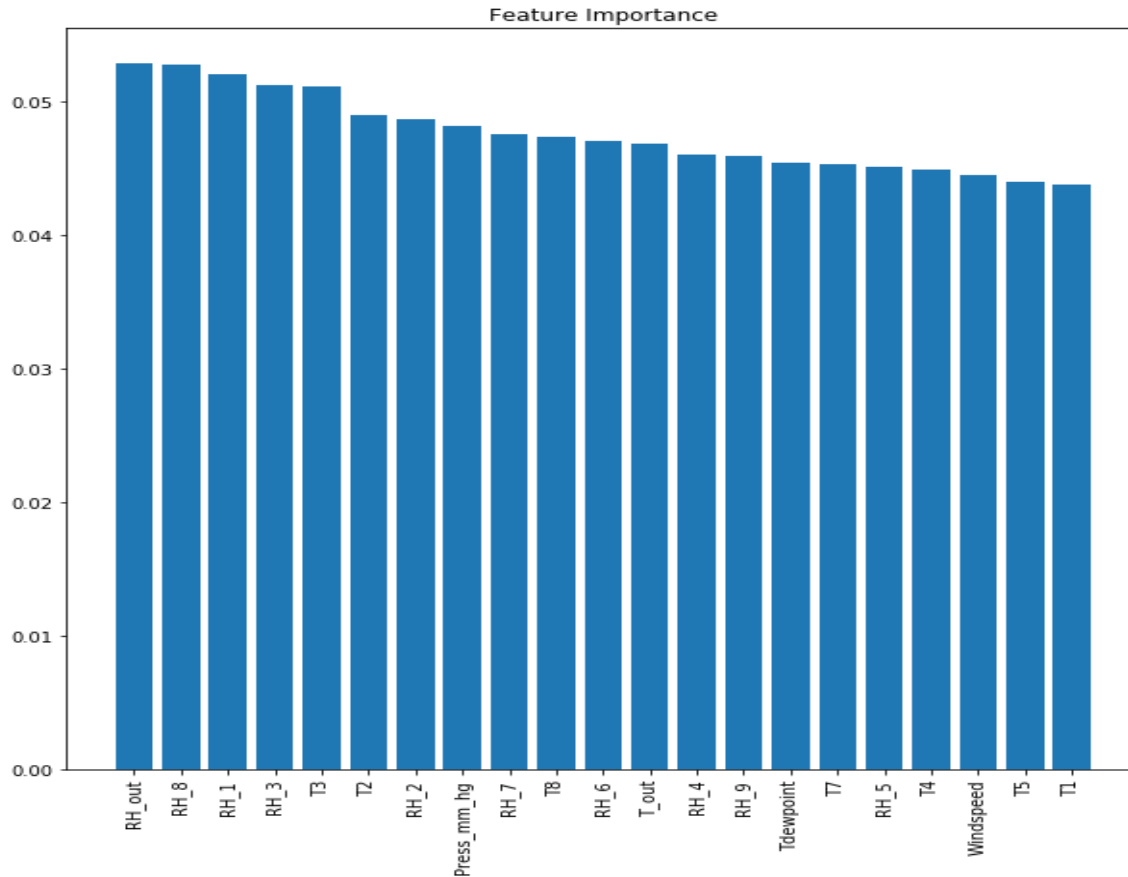
Вхідні дані дата сету

Діаграма кореляції даних



- Дата: Рік, місяць, день, година, хвилина та секунда.
- Appliances: Споживання енергії у Wh.
- lights: Енерговикористання світильників у будинку в Wh .
- T1: Температура в кухонній зоні, в Цельсіях.
- RH_1: Вологість кухонних приміщень,%.
- T2: Температура в зоні вітальні, в Цельсіях.
- RH_2: Вологість в площі вітальні, у%.
- T3: Температура в площі пральні.
- RH_3: Вологість в площі пральні, в%.
- T4: Температура в офісному приміщенні, в Цельсіях.
- RH_4: Вологість в офісному приміщенні , в%.
- T5: Температура у ванній кімнаті, в Цельсіях.
- RH_5: Вологість у ванній кімнаті, у%.
- T6: Температура поза будівлею (північна сторона), в Цельсіях.
- RH_6: Вологість біля будівлі (північна сторона), у%.
- T7: Температура в прасуванні кімната, у Цельсіях.
- RH_7: Вологість у прасувальній кімнаті, у%.
- T8: Температура в кімнаті для підлітків 2, у Цельсіях.
- RH_8: Вологість у кімнаті для підлітків 2, у%.
- T9: Температура в кімнаті для батьків, в Цельсіях.
- RH_9: Вологість в кімнаті для батьків, у%.
- T_out: Температура зовні (від метеостанції Chievres).
- Press_mm_hg: Тиск Цельсія (від метеостанції в Chievres), мм мм рт.ст.
- RH_out: Вологість біля будівлі у%.
- Windspeed: Швидкість вітру (від метеостанції в Chievres) в м / с.
- Visibility: Видимість (від метеостанції Chievres), в км.
- Tdewpoint: (від метеостанції Chievres), ° C.
- rv1: Випадкова змінна 1.
- rv2: Випадкова змінна 2.

Діаграма важливості атрибутів



Обрані методи для вдосконалення

```
{'Name': 'KNeighborsRegressor: ',  
  'Train_Time': 0.05581855773925781,  
  'Train_R2_Score': 0.6814639376039895,  
  'Test_R2_Score': 0.4855598647200512,  
  'Test_RMSE_Score': 0.7172448224141802},
```

```
{'Name': 'RandomForest ',  
  'Train_Time': 3.095679998397827,  
  'Train_R2_Score': 0.9136912792970885,  
  'Test_R2_Score': 0.5256812270127812,  
  'Test_RMSE_Score': 0.6887080462628695},
```

```
{'Name': 'ExtraTreeRegressor: ',  
  'Train_Time': 0.7568228244781494,  
  'Train_R2_Score': 1.0,  
  'Test_R2_Score': 0.5771832228819963,  
  'Test_RMSE_Score': 0.6502436290483773},
```

ExtraTreesRegressor

Алгоритм Extra-Trees створює ансамбль необрізаних дерев рішень або регресії згідно класичної процедури зверху вниз. Дві його основні відмінності від інших методів, що базуються на деревах, полягають у тому, що він розділяє вузли, вибираючи вирізані точки навмання, і що для вирощування дерев використовує цілу навчальну вибірку.

RandomForest

це ансамблевий метод, який тренує кілька дерев рішень паралельно із завантаженням з подальшим агрегуванням. Початкове завантаження вказує на те, що кілька окремих дерев рішень навчаються паралельно на різних підмножинах набору навчальних даних, використовуючи різні підмножини доступних функцій.

k -найближчих сусідів

метричний алгоритм для автоматичної класифікації об'єктів. Основним принципом методу найближчих сусідів є те, що об'єкт присвоюється тому класу, який є найбільш поширеним серед сусідів даного елемента. Сусіди беруться, виходячи з множини об'єктів, класи яких уже відомі, і, виходячи з ключового для даного методу значення k , враховується, який клас є найчисленнішим серед них

Результати прогнозу після вдосконалення

ДО

```
'Name': 'ExtraTreeRegressor ',  
'Train_Time': 0.74996018409729,  
'Train_R2_Score': 1.0,  
'Test_R2_Score': 0.5771832228819963,  
'Test_RMSE_Score': 0.6502436290483773},
```

```
{'Name': 'RandomForest ',  
'Train_Time': 3.095679998397827,  
'Train_R2_Score': 0.9136912792970885,  
'Test_R2_Score': 0.5256812270127812,  
'Test_RMSE_Score': 0.6887080462628695},
```

```
{'Name': 'KNeighborsRegressor: ',  
'Train_Time': 0.05581855773925781,  
'Train_R2_Score': 0.6814639376039895,  
'Test_R2_Score': 0.4855598647200512,  
'Test_RMSE_Score': 0.7172448224141802},
```

ПІСЛЯ

```
Training set R2 Score - 100.0  
Testing set R2 Score - 68.48834662363605  
Testing set RMSE Score - 60.42487350120308
```

```
Training set rfr R2 Score - 67.73102771588601  
Testing set rfr R2 Score - 41.48186453266495  
Testing set rfr RMSE Score - 76.49714731108281
```

```
Training set knr R2 Score - 73.23026930811155  
Testing set knr R2 Score - 57.52448943875144  
Testing set knr RMSE Score - 65.17323880339887
```

Наукова новизна

Основні результати, які були отримані в процесі вирішення поставлених завдань та становлять наукову новизну дослідження, полягають у вдосконаленні інформаційної технології прогнозування енергоспоживання електрообладнання будівлі, що дозволяє знизити енергоспоживання за рахунок технології інтернету речей.

Практичне значення одержаних результатів полягає у можливості користувача даної інформаційної технології:

- прогнозувати показники енергоспоживання та ефективно розподіляти роботу електрообладнання будівлі з низьким енергоспоживанням;
- підбирати кращі тарифні плани серед тих, що пропонують електроенергетичні компанії;
- оцінити рівень енергоефективності будівлі з урахуванням показників зовнішнього середовища, отриманого за допомогою інтернету речей.

Висновки

Обґрунтовано доцільність вдосконалення технології та важливість оптимального використання енергоресурсів.

Проаналізовано схожі рішення.

Виконано розвідувальний аналіз даних – нульові значення та ті, які не мали сильного зв'язку з вихідною величиною, було видалено з вибірки.

Було проведено нормалізацію даних, виконано порівняння з іншими моделями та методами прогнозування та обрано кращий (kNN).

Вдосконалено ключову прогнозну модель інформаційної технології.

Дякую за увагу!