

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

МОДЕЛЬ ВИЯВЛЕННЯ АНОМАЛІЙ У ДАНИХ ПРО СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

Автор:
студент групи СПм-22-5
Братищенко М.Р.

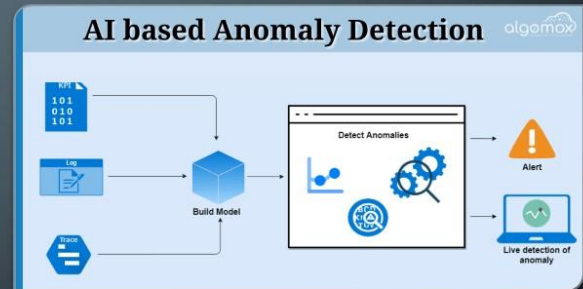
Керівник:
к.т.н., доцент каф. ЕОМ
Філімончук Т.В.

МЕТА:

- дослідження існуючих рішень щодо виявлення аномалій в даних;
- застосування передових технік та технологій для ідентифікації нетипових значень в даних про енергоспоживання;
- запропонування власного рішення проблеми виявлення аномалій в показниках електроспоживання.

ІСНУЮЧІ МЕТОДИ ВИЯВЛЕННЯ АНОМАЛІЙ

- Isolation Forest
- Clusterization of K-means
- Support Vector Machine
- Local Outlier Factor
- Unsupervised Niche Clustering
- Vectored Autoregression (VAR)
- Autoregressive integrated moving average (ARIMA)



3

ІСНУЮЧА МОДЕЛЬ ВИЯВЛЕННЯ АНОМАЛІЙ У ДАНИХ ПРО СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

$ADM-ECD = \{ВДЕ, АМНВА, ВР\}$

- ВДЕ – вхідні дані щодо енергоспоживання;
- АМНВА – алгоритм машинного навчання для виявлення аномалій;
- ВР – відображення результатів.

4

МОДЕЛЬ CADM-ECD, ЩО ЗАПРОПОНОВАНА В РОБОТІ

CADM-ECD = {ВДЕ, МВА, МГВ}

МВА = {ППМАКС, ПРАІЛ }

- ВДЕ – вхідні дані щодо енергоспоживання;
- МВА – модуль виявлення аномалій;
- МГВ – модуль графічного відображення.
- ППМАКС – підмодуль побудови моделі авторегресії ковзного середнього (ARIMA);
- ПРАІЛ – підмодуль з реалізацією алгоритму «Ізоляційний ліс».

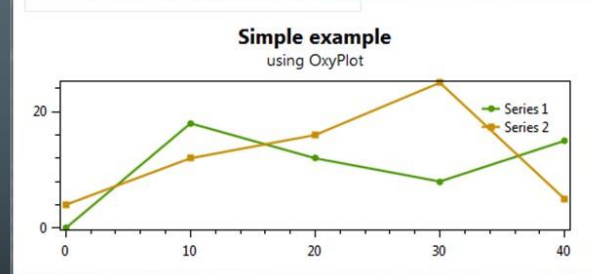
5

АЛГОРИТМ РОБОТИ МОДЕЛІ CADM-ECD

- Крок 1: завантаження даних про споживання електроенергії.
- Крок 2: розділення даних на навчальний та тестувальний набори.
- Крок 3: побудова моделі ARIMA для прогнозування споживання електроенергії.
- Крок 4: прогнозування значень споживання електроенергії за допомогою побудованої ARIMA моделі.
- Крок 5: визначення залишків між фактичними та прогнозованими значеннями.
- Крок 6: використання Isolation Forest для виявлення аномальних залишків.
- Крок 7: виявлення аномальних точок на основі оцінки Isolation Forest.
- Крок 8: відображення графіка з фактичними даними, прогнозом, залишками та виявленими аномаліями.

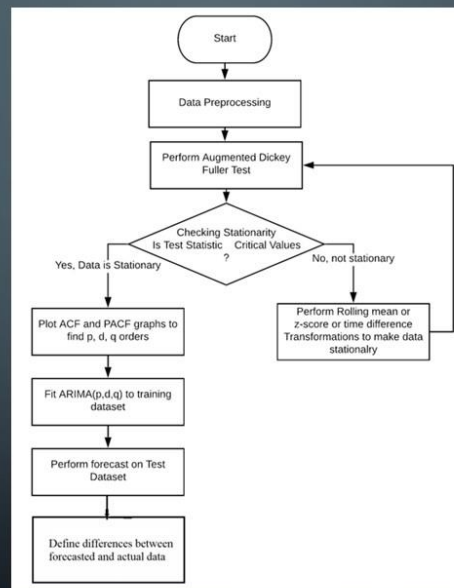
6

ВИБІР ІНСТРУМЕНТІВ ТА ТЕХНОЛОГІЙ ДЛЯ РЕАЛІЗАЦІЇ МОДЕЛІ САДМ-ЕСД



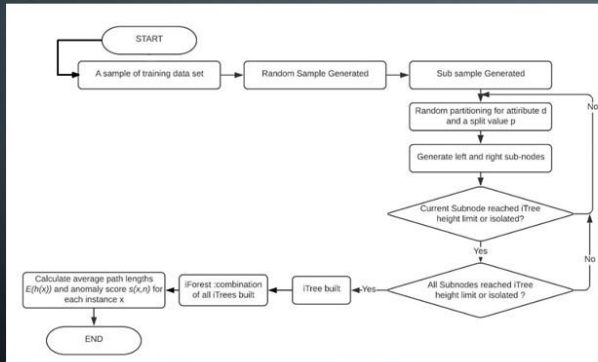
7

ПРИНЦИП РОБОТИ ПІДМОДУЛЯ ППМАКС

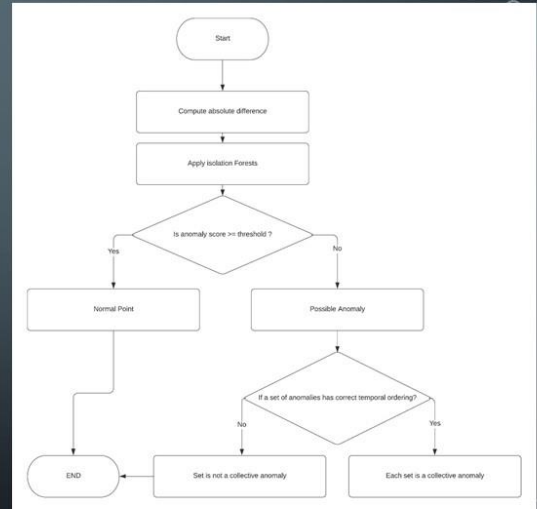


8

ПРИНЦИП РОБОТИ ПІДМОДУЛЯ ПРАІА



9



ТЕСТУВАННЯ МОДЕЛІ САДМ-ЕСД

Date and time Energy Consumption (kWh)

...

2022-01-05 02:00:00 202.538242

2022-01-05 03:00:00 208.039787

2022-01-05 04:00:00 450.000000

2022-01-05 05:00:00 173.044819

...

2022-01-09 07:00:00 226.730559

2022-01-09 08:00:00 30.000000

2022-01-09 09:00:00 195.212416

2022-01-09 10:00:00 221.993192

...

2022-01-21 18:00:00 179.639162

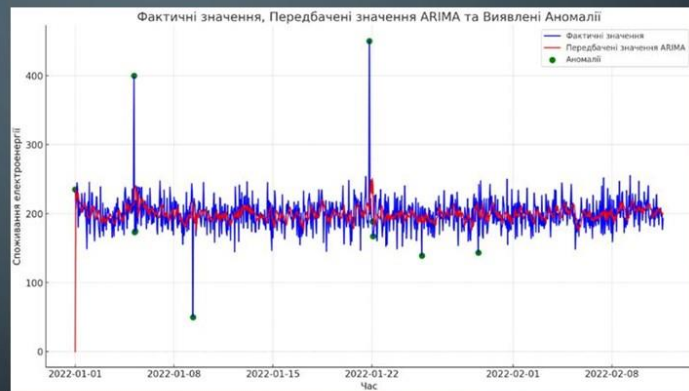
2022-01-21 19:00:00 198.442905

2022-01-21 20:00:00 450.000000

2022-01-21 21:00:00 199.315154

2022-01-21 22:00:00 221.926937

...



10

ОБ'ЄКТИВНІ ТА СУБ'ЄКТИВНІ ОЦІНКИ МОДЕЛІ CADM-ECD

Обчислені метрики:

- Accuracy = 0.993;
- ROC-AUC = 0.9965;
- Precision = 0.3;
- Recall = 1.0;
- F1-Score = 0.4615.

Переваги:

- точність, надійність;
- легкість розуміння результатів;
- адаптація до різних частот даних.

Недоліки:

- менша ефективність з нелінійними або сильно змінними даними;
- вимагання значних обчислювальних ресурсів та ретельне налаштування для великих обсягів даних.

11

ВИСНОВКИ



- ефективність моделі CADM-ECD;
- потенціал для реального застосування;
- вагоме значення для сфери електроенергетики;
- зменшення споживання електроенергії;
- підвищення енергоефективності;
- перспективи розвитку запропонованої моделі.

12