



ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

01

Слайди go go no вігі

Виконав:
студент МП-2
"Комп'ютерні Науки"
Медведев Д.Р.

Науковий керівник:
Жежерун О.П.

Проблема

ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН
АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У
БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

02

Labels	Timestamp	Util	Temp
name="CPU", host="host_1", cpu_id="1"	1574260177	0.95	48
name="CPU", host="host_1", cpu_id="1"	1574260216	0.83	32
name="CPU", host="host_1", cpu_id="1"	1574260325	0.93	45
name="CPU", host="host_1", cpu_id="1"	1574260422	0.76	42
name="CPU", host="host_2", v_cpu_id="1"	1574260163	0.03	20
name="CPU", host="host_2", v_cpu_id="1"	1574260173	0.00	19
name="CPU", host="host_2", v_cpu_id="1"	1574260183	0.01	18
name="CPU", host="host_2", v_cpu_id="1"	1574260193	0.05	21
name="CPU", host="lab_1", cpu_id="0", cluster="c_0"	1574260002	1.00	60
name="CPU", host="lab_1", cpu_id="0", cluster="c_0"	1574260012	1.00	60
name="CPU", host="lab_1", cpu_id="0", cluster="c_0"	1574260035	1.00	60

RCA Memogu

ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН
АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У
БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

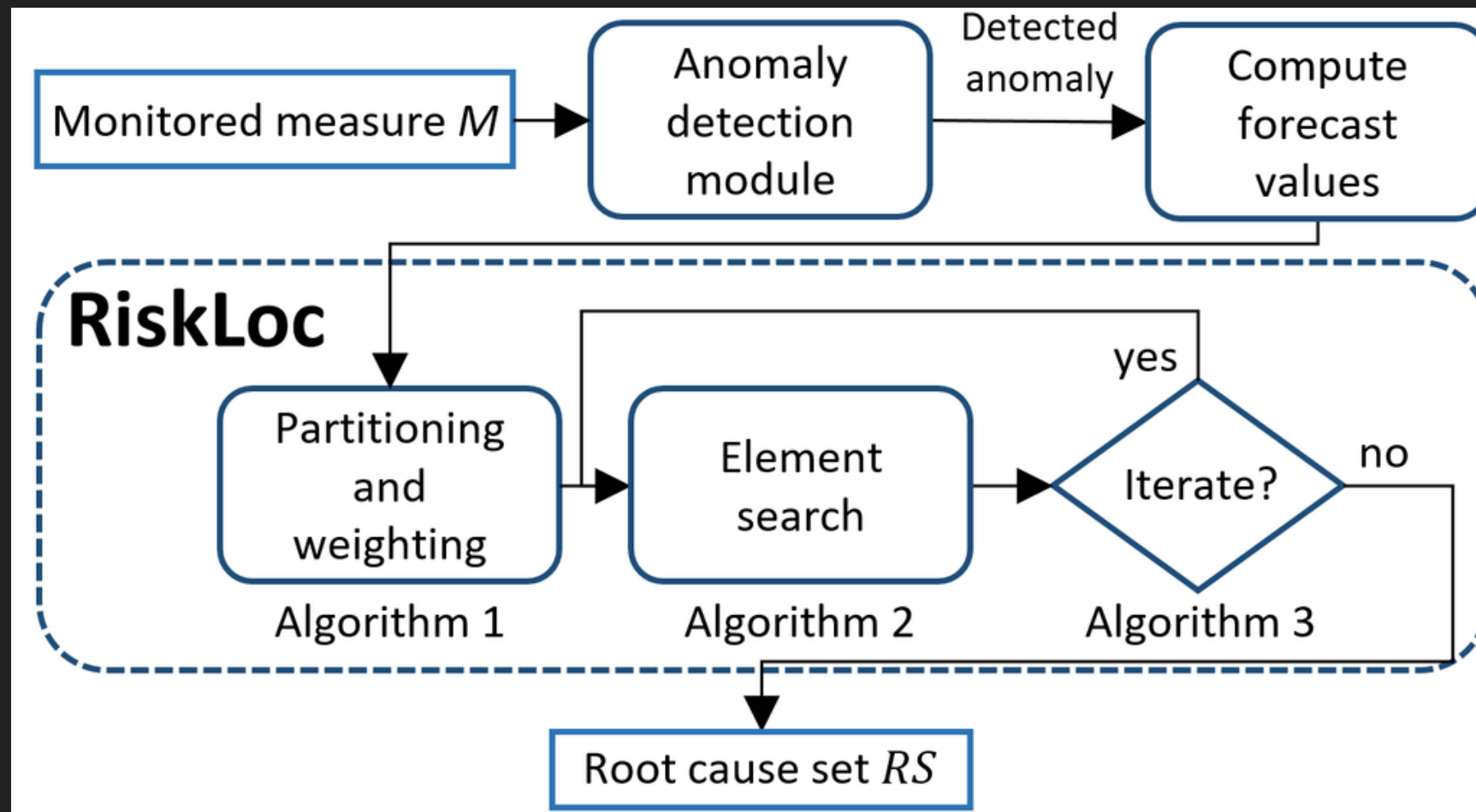
03

Memog (Pik)	Багатовимірна підтримка	Множинні першопричини	Залежність від прогнозів	Ключова інновація
Adtributor (2014)	Ні — одне вимірювання	Ні	Так — очікувані значення	Jensen-Shannon divergence для відхилень
iDice (2016)	Так — комбінації значень	Так	Так — вимірювання відхилень	Статична відсічка комбінацій за значимістю
MacroBase (2017)	Так — асоціативна аналітика	Ні	Ні — асоціативні правила	Асортимент асоціативних патернів у логах
Squeeze (2019)	Так — групування сегментів	Так	Так — ripple-effect score	Підхід «спершу групуй, потім ранжуй»
MID (2020)	Так — еволюційний пошук	Так	Так — оцінка fitness	Генетичний алгоритм для пошуку сегментів
AutoRoot (2021)	Так — адаптивна кластеризація	Так	Так — KDE для кластерів	KDE для автоматичного підбору кластерів
HALO (2021)	Так — ієрархічне занурення	Так	Так — очікує ієрархії	Двоетапний пошук із ієрархічними рівнями
RiskLoc (2022)	Так — рекурсивне розбиття	Так	Так — базові прогнози	Ризик-зважене розбиття з ripple-effect

RiskLoc (2022)

ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН
АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У
БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

04



а) Залежність від точних прогнозів сегментів ускладнює локалізацію першопричин за обмежених історичних даних.

б) Експоненціальне розбиття сегментів призводить до надмірної деталізації та зростання обчислювальних витрат часу.

05

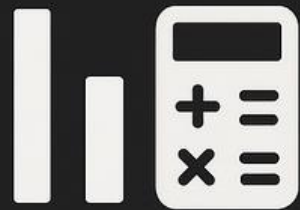
Мета

Розробити та оцінити
forecast-free аномалію RiskLoc
для ефективної локалізації
багатовимірних першопричин
метрик.

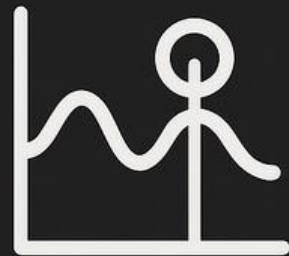
Методологія

ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН
АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У
БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

06



BASELINE CALCULATION



NORMALIZATION & ANOMALY SCORING



CLUSTERING & RISKLOC

Базові статистичні очікування

Замінили прогнози на історичне середнє, медіану, peer-median, parent-sum і зважені комбінації для кожного сегмента. Це дозволило оцінювати відхилення без залежності від зовнішніх моделей.

Нормалізація та оцінка аномалій

Використали z-score для уніфікації відхилень і anomaly degree для пріоритизації сегментів із найгучнішими аномаліями. Це підвищило стійкість до різних масштабів та шумів у даних.

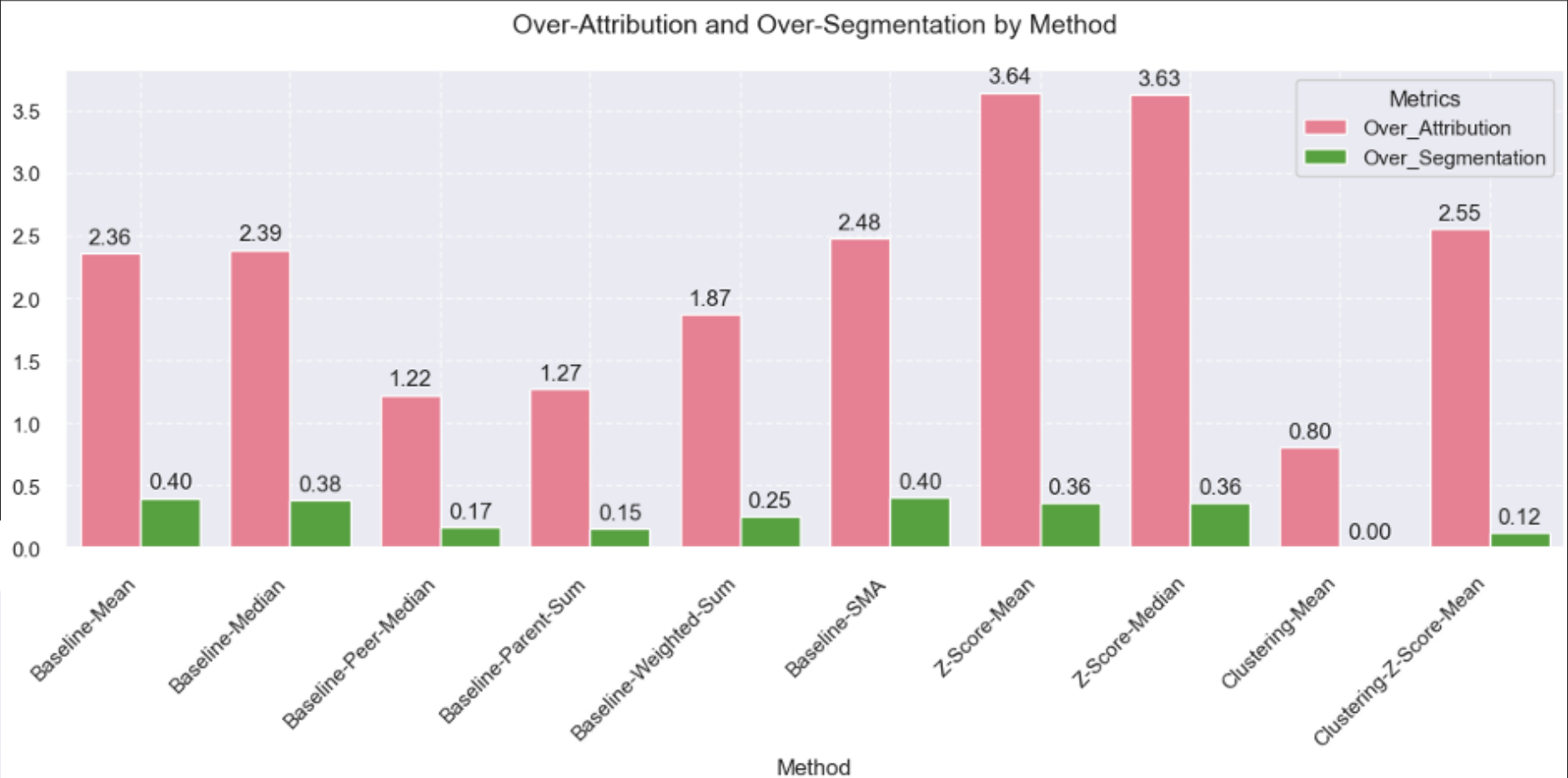
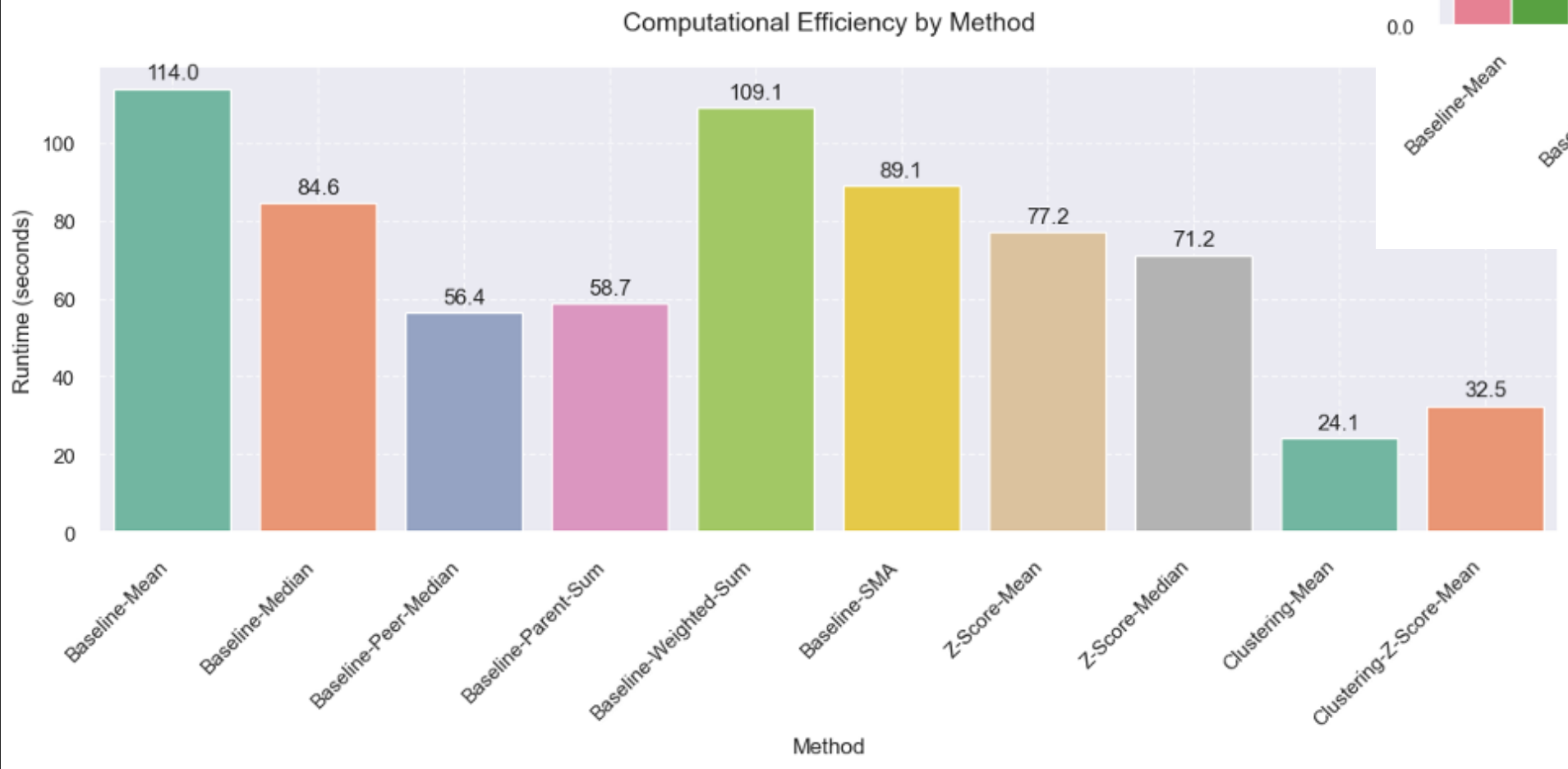
Кластеризація сегментів

Застосували k-means до відхилень для групування схожих аномальних сегментів і запустили RiskLoc окремо в кожному кластері. Такий підхід зменшив надмірну атрибуцію та скоротив обчислювальні витрати.

Результати (ч1)

ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН
АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У
БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

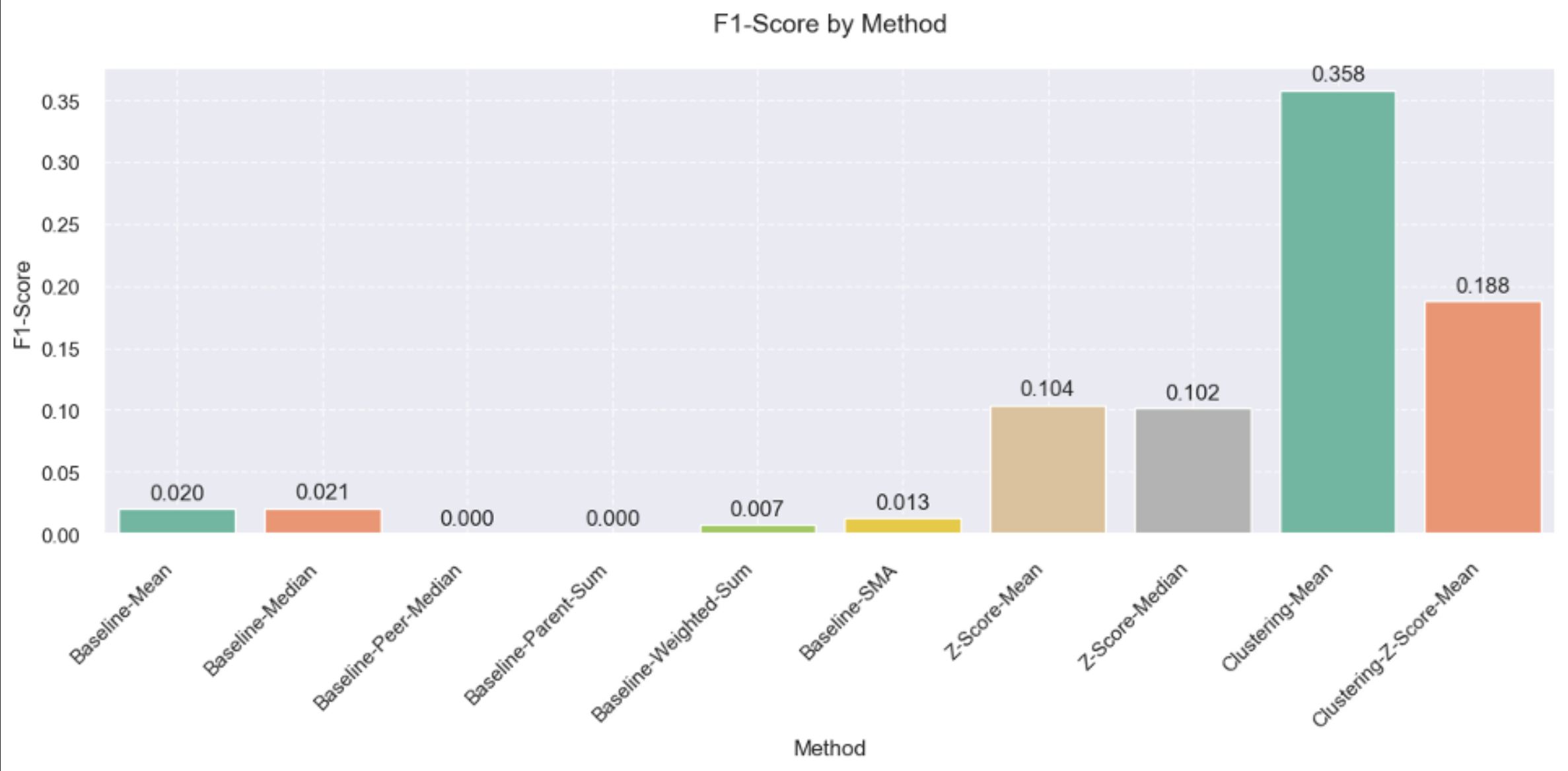
07



Результати (ч2)

ПРОГНОЗ-НЕЗАЛЕЖНА ЛОКАЛІЗАЦІЯ ПЕРШОПРИЧИН
АНОМАЛІЙ З КЛАСТЕРИЗАЦІЄЮ СЕГМЕНТІВ У
БАГАТОВИМІРНИХ ЧАСОВИХ РЯДАХ

08



Висновки

Anomaly Detection & RCA

Проведено системний огляд сучасних підходів до виявлення аномалій та локалізації першопричин (RCA). Виокремлено ключові виклики: залежність від прогнозів, надмірна деталізація, масштабованість

RiskLoc Adoption

Застосовано forecast-free налаштування, anomaly degree. Інтегровано кластеризацію k-means для попереднього відсіювання шумових сегментів. Оновлено генератор синтетичних датасетів.

Experiments & Analysis

Проведено серію експериментів на наборі RS. Найкраща продуктивність у Clustering-Mean ($F1 = 0.36$, runtime = 24 c), суттєве покращення порівняно з базовими методами.

Дякую за увагу