

RedLab/Hack

Разработка модели для выявления аномалий
во временном ряду

Команда: `ikanam_chipi_chipi`

Проблематика



Скорость работы

Требуется алгоритм, который **быстро определяет** аномалии в данных



Open-source

Необходимо использовать только **бесплатные, открытые решения**



Сезонность

Внутри дня метрики могут меняться значительно.

Изменения вызванные сезонностью – не аномалии



Аномалия - это что?

Необходимо определить, что считается **«ненормальным»** поведением метрик



Наше решение

Интерфейс, который позволяет анализировать временной ряд и отмечать выявленные аномалии в данных.

Доступные опции:

- Загрузка .tsv файла
- Выбор временного интервала для анализа
- Интерактивные графики
- Выгрузка файла с отмеченными аномалиями

RedLab Hack

Team: ikanam_chipi_chipi

Enter the path to your metrics_collectors file

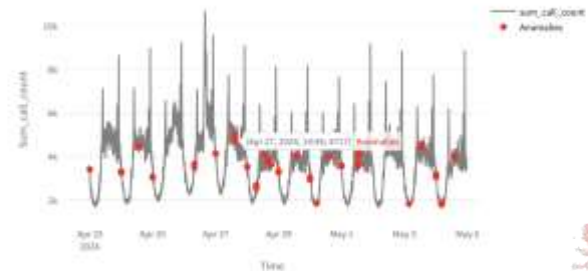
metrics_collectors.tsv

Process

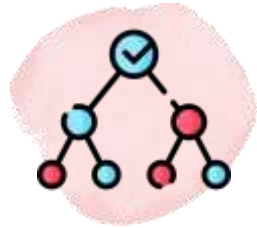
Select time interval

2024-04-22 2024-05-04
2024-04-15 2024-05-16

Throughput Anomalies



Стек



Isolation Forest

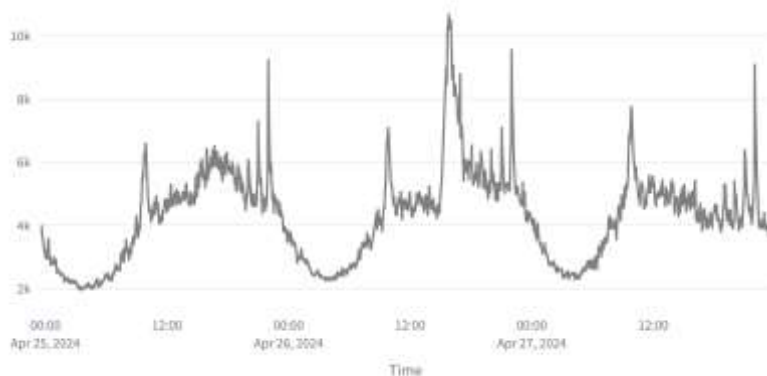
Алгоритм машинного обучения для поиска аномалий с учётом нескольких признаков одновременно



Streamlit

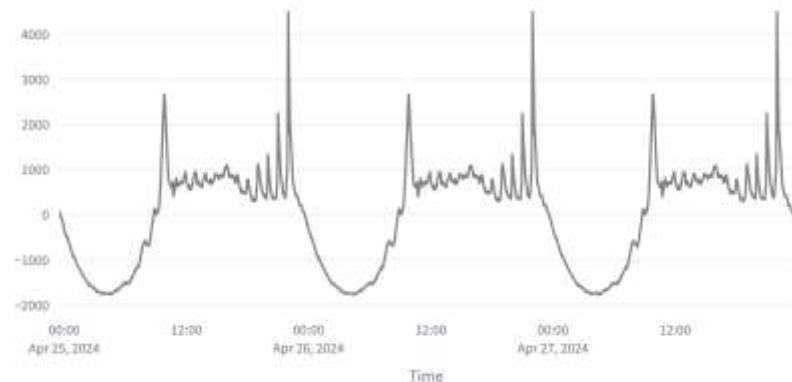
Python библиотека для создания веб-интерфейса с возможностью добавления интерактивных графиков

Борьба с внутридневной сезонностью



Метрика до обработки

Не всегда понятно – большое значение связано со временем суток или же это аномалия?



Компонента сезонности

Используя библиотеку statsmodels, находим **компоненту сезонности**

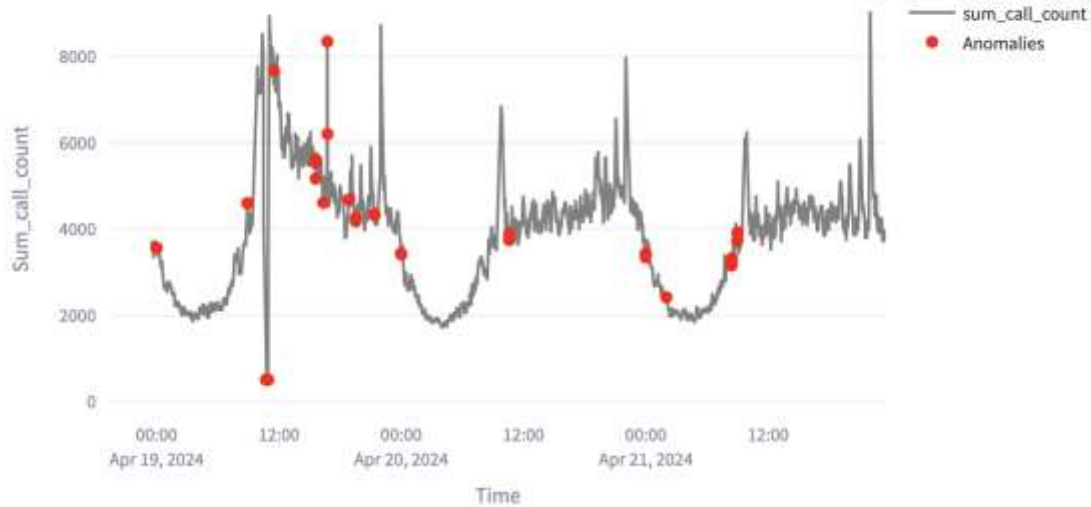
Отчетливо видна **аномалия!**



Удалили компоненту сезонности

Не аномалии

Throughput Anomalies



метрика до обработки

Уникальность решения

Благодаря тому, что мы учли сезонность,
алгоритм удалил экстремальные значения не «в тупую»

Алгоритм формирует **anomaly_score** –
вероятность, что объект является
аномалией.

Можно задать свой трешхолд.

	point	seasonally_adjusted_sum_call_count	seasonally_adjusted_web_response	seasonally_adjusted_apdex	seasonally_adjusted_error_rate	anomaly	anomaly_score
19,531	2024-04-29 13:15:00	3,458.9023	0.0146	0.9958	0.002	<input type="checkbox"/>	0.38
19,532	2024-04-29 13:16:00	3,387.0818	0.1488	0.9946	0.0028	<input checked="" type="checkbox"/>	0.78
19,533	2024-04-29 13:17:00	3,392.3141	0.1535	0.9946	0.0026	<input checked="" type="checkbox"/>	0.78
19,534	2024-04-29 13:18:00	3,405.3919	0.0185	0.9975	0.0003	<input type="checkbox"/>	0.08

Уникальность решения

Если увеличить трешхолд – с наибольшей вероятностью
будут указаны только аномалии.

Уменьшить – все подозрения на аномалии будут
отображены.

Интерпретируемость решения

☒ Explanation of model decision

Choose exact timestamp to get explanation of model's decision:

2024-04-22 10:11:00



**Есть возможность получить ответ на
вопрос:**

*Руководствуясь какими факторами модель
определила объект как аномалию?*

Демонстрация решения

https://github.com/maxlyara1/find_anomalies_hackathon/blob/main/%D0%98%D0%BD%D1%82%D0%B5%D1%80%D1%84%D0%B5%D0%B9%D1%81_26_05_2024.mp4

Масштабируемость решения



Подключение к БД

Для использования на практике, в реальном времени есть возможность подключиться к базе данных для **прямого доступа к данным**



Добавление метрик

Есть возможность дополнить существующие метрики для **увеличения вероятности обнаружения сбоя в работе системы** есть

ikanam_chipi_chipi



Аделя Сабирова
Data engineer, designer



Максим Ляра
Team lead, Data scientist
TG: @maxlyara1



Станислав Палатов
Data scientist