

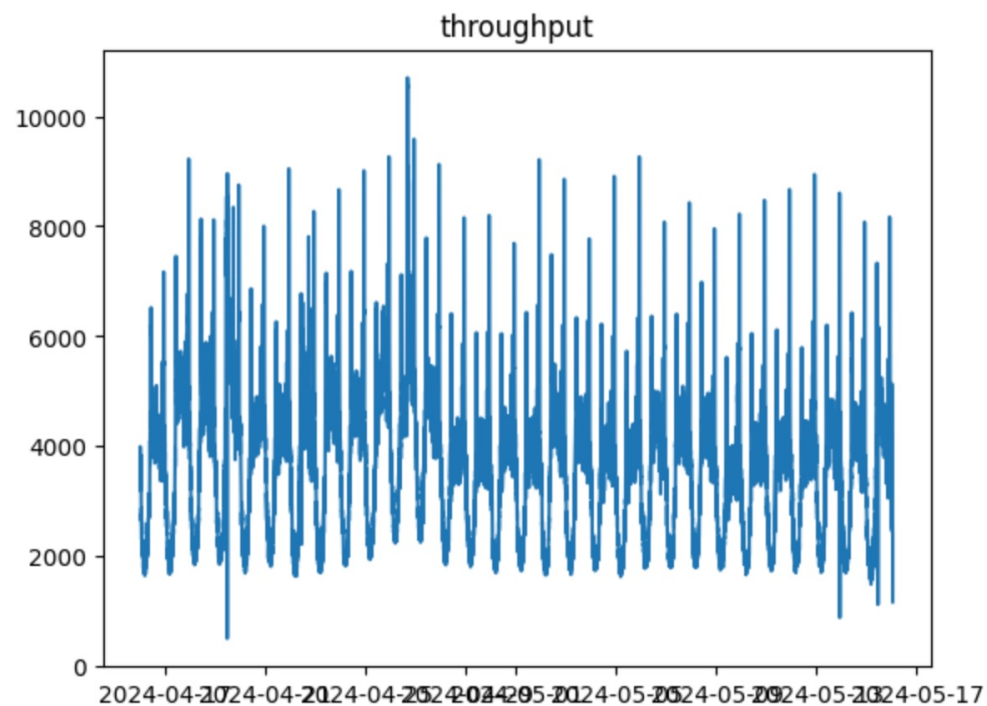
Red Lab Hack(Case 1)

Команда ETNA

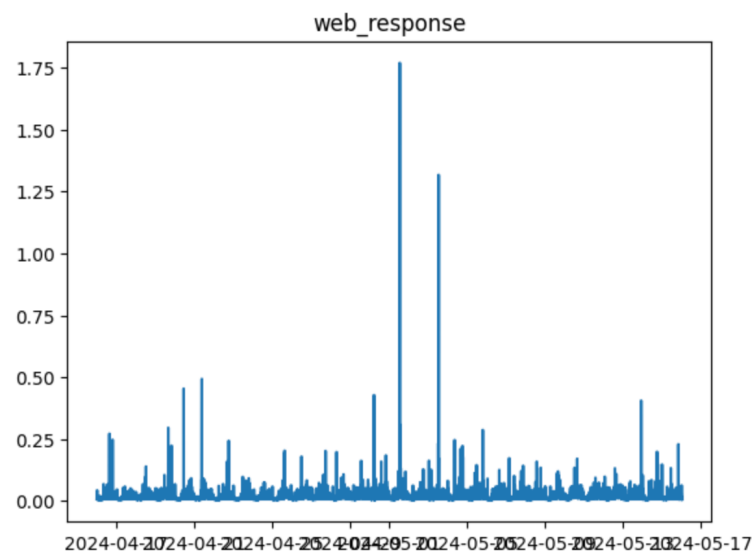
План доклада

- EDA
- Оценка качества
- Алгоритм детекции
- Сервис
- Преимущества

EDA



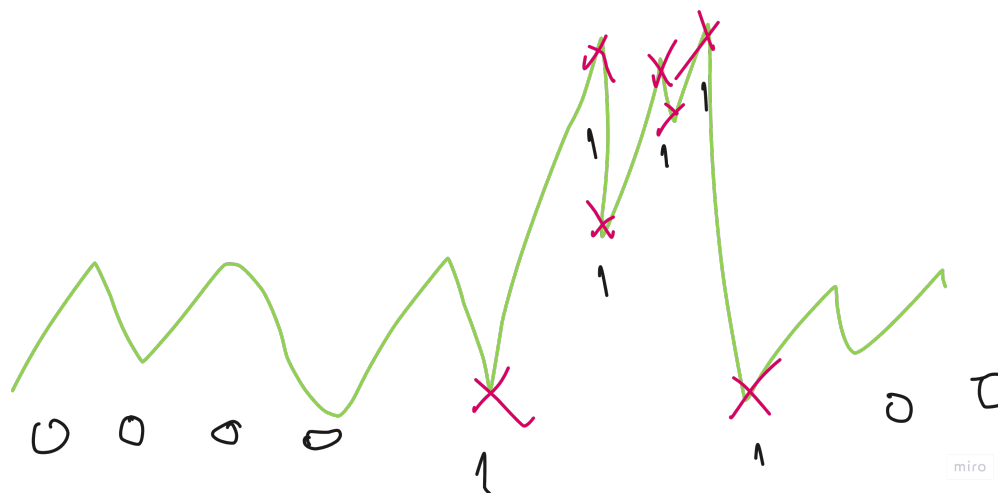
1. **Пропуски** → ffill (состояние ряда сохраняется во время пропусков + не подсматриваем в будущее)
2. **Сезонность** → важно детектировать и использовать в моделях



Постановка задачи

$$[t_1, \dots, t_n] \rightarrow [0, 1, 1, \dots, 1, 0]$$

Отрезок Метка аномальности

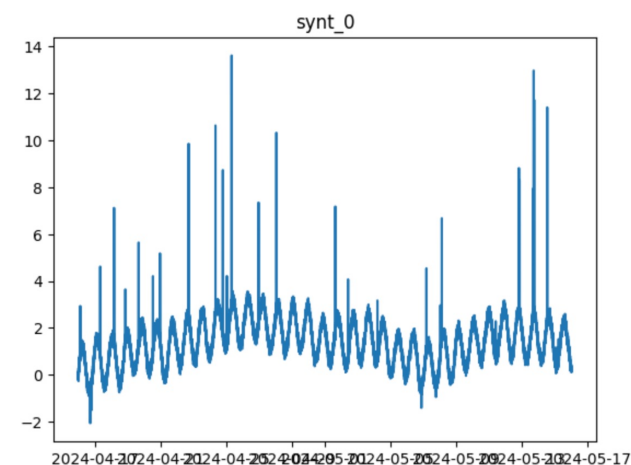
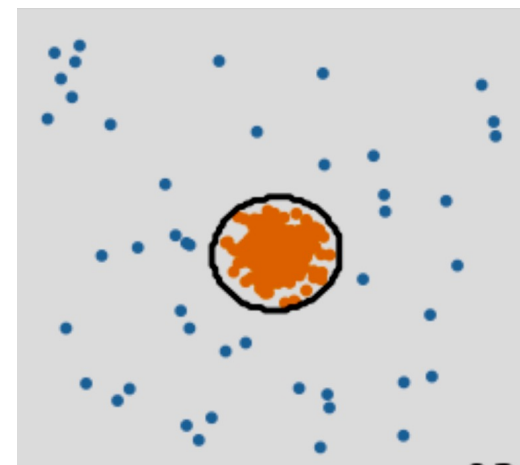


Постановка задачи

Оценка качества(метрики кластеризации):

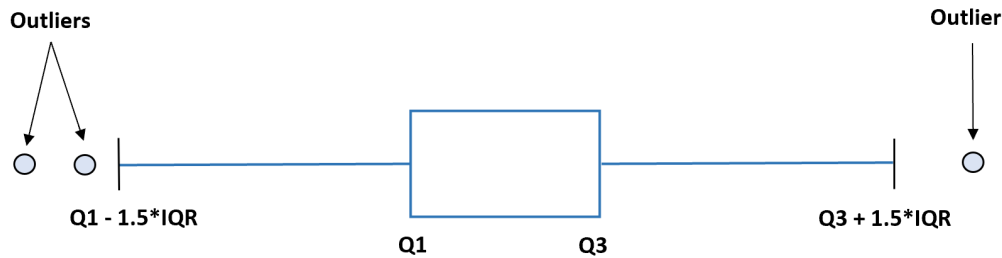
$A * \text{Hausdorff}(\text{модификация}) + B * \text{Silhouett} + C * \text{Davies_Bouldin} \rightarrow \text{MAX}$

- Вариант №1: Отсутствие разметки \rightarrow Unsupervised метрик
- Вариант №2:
 - Сделать ручную разметку(при помощи наивного алгоритма)
 - Supervised метрики классификации(Recall, Precision, F-beta)
- Вариант №3:
 - Сгенерировать синтетические данные с аномалиями
 - Проверять качество на известной разметке



Алгоритмы

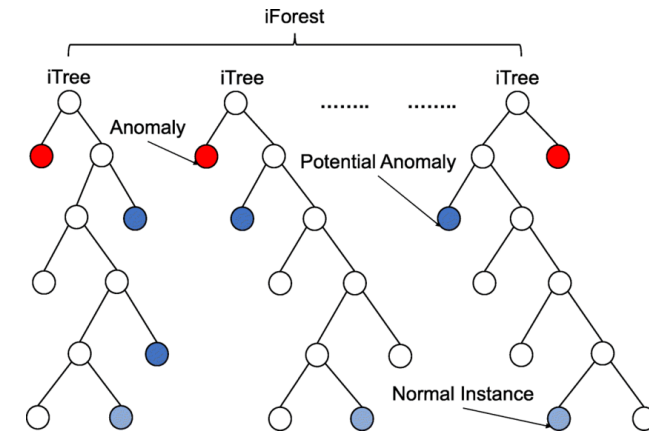
IQR



Особенности:

- Динамика: оценка на скользящих окнах
- Сезонность и тренд: STL разложение
- Оценка уверенности

Isolation Forest

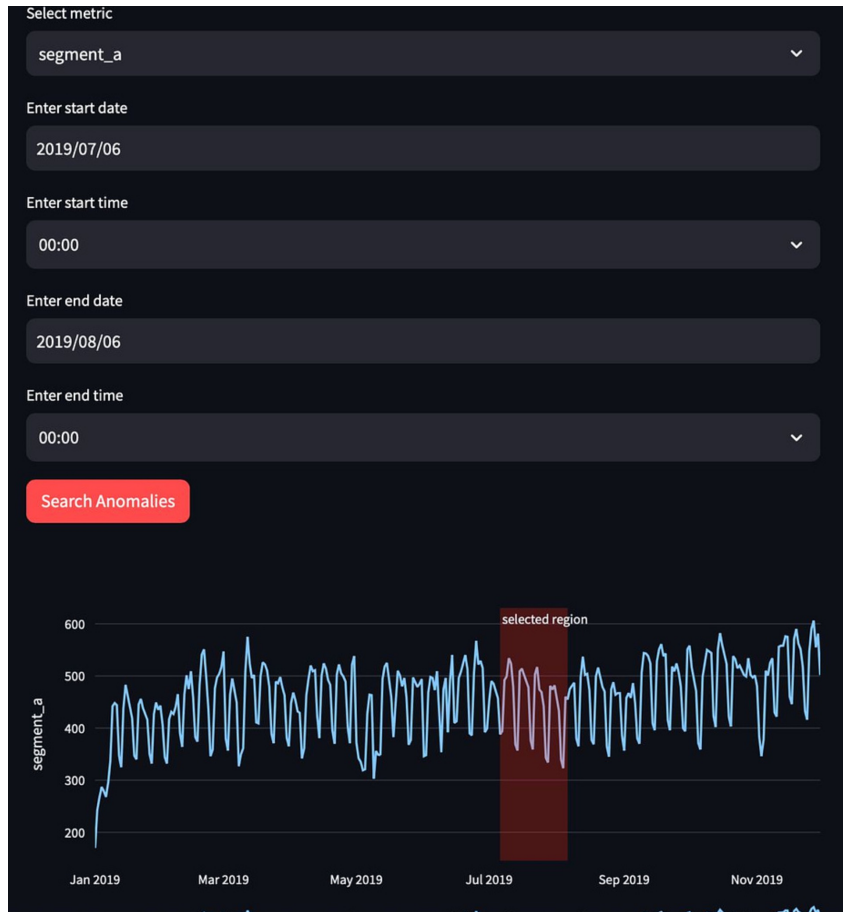


Особенности:

- Признаковое представление
 - Компоненты из Holt-Winters
 - Признаки из временной метки
- Быстрее IQR
- Оценка уверенности

+ Автоматическое определение сезонности через периодограммы

Сервис



Деплой



Веб-интерфейс



Данные

Преимущества решения

1. Не требует разметки
2. Комплексная метрика оценки качества
3. Легковесная модель
4. Оценка сора аномальности для каждого наблюдения
5. Модели учитывают сезонность и тренд
6. Автоматический расчёт периодичности сезонности

Спасибо за внимание!