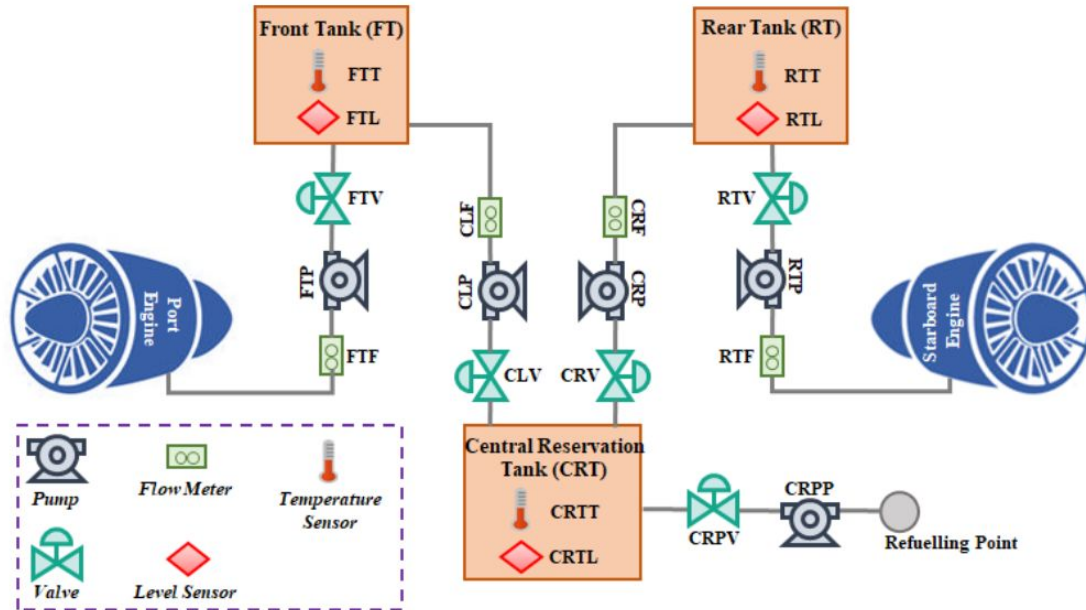


Метод обнаружения аномальных режимов работы двигателей системы впрыска топлива

Осенний семестр, 2023-2024

Система впрыска топлива двигателя самолета



Pipeline

- Сборка датасета.
- Обучение модели прогнозирования (Linear Progression, Random Forest).
- Обучение модели обнаружения аномалий.
- Распознавание аномалий, вывод метрик (NormScorer, KMeansScorer).

Сборка датасета

Будем расширять датасет из периодов нормального и аномальных сценариев работы системы.

- 1) Создадим временной ряд из 500 периодов нормального сценария.
 - a) $3/4$ поместим в обучающий датасет
 - b) Оставшуюся $1/4$ – в тестовый.
- 2) Создадим временной ряд из 5 периодов различных аномальных сценариев.
 - a) Далее эти периоды будут произвольным образом вставляться в тестовый набор данных.

Таким образом, обучающий датасет не содержит аномалий, в отличие от тестового.

Augmenter

```
augmenter = (  
    TimeWarp(n_speed_change=2, max_speed_ratio=1.2)  
    + Dropout(p=0.05, fill='ffill')  
    + AddNoise(scale=0.005)  
)
```

1. Искажение времени (TimeWarping) с 2 изменениями скорости, максимальное соотношение максимальной и минимальной скорости было установлено равным 1,2.
2. Из временного ряда значение каждой точки ряда заменяется “левым соседом” с вероятностью 5%.
3. К каждой точке добавляется шум – независимая случайная величина нормального распределения с нулевым средним и стандартным отклонением 0,005.

Прогнозирование

1. Линейная регрессия

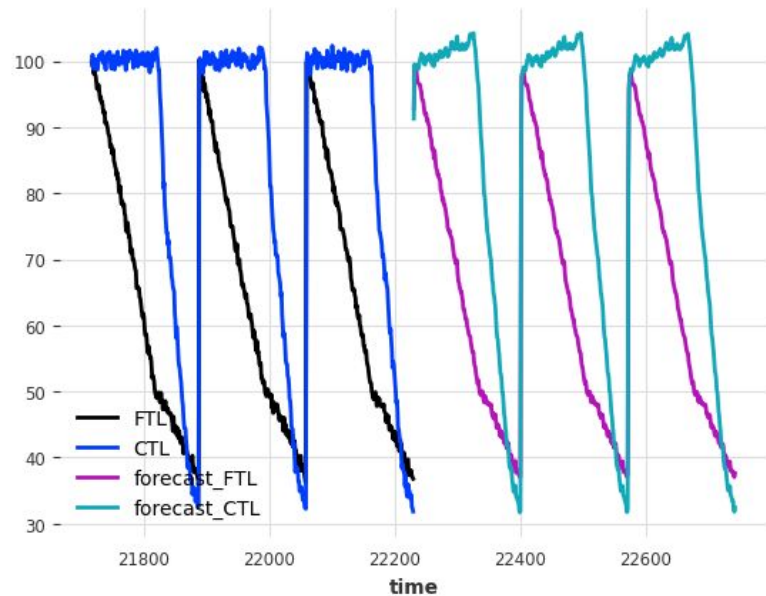
На основе данных фиксированного размера (в нашем случае, периода) до произвольного момента времени расширяет данные на некоторый промежуток времени дальше.

2. Метод случайного леса

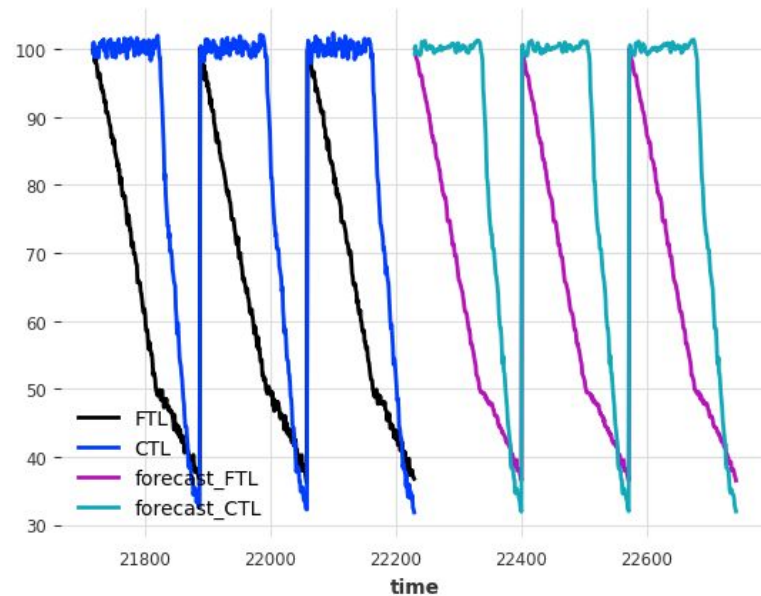
Большое количество деревьев решений строится с использованием случайно выбранных подмножеств обучающих данных и признаков.

Прогнозирование: отличия

LinearRegression

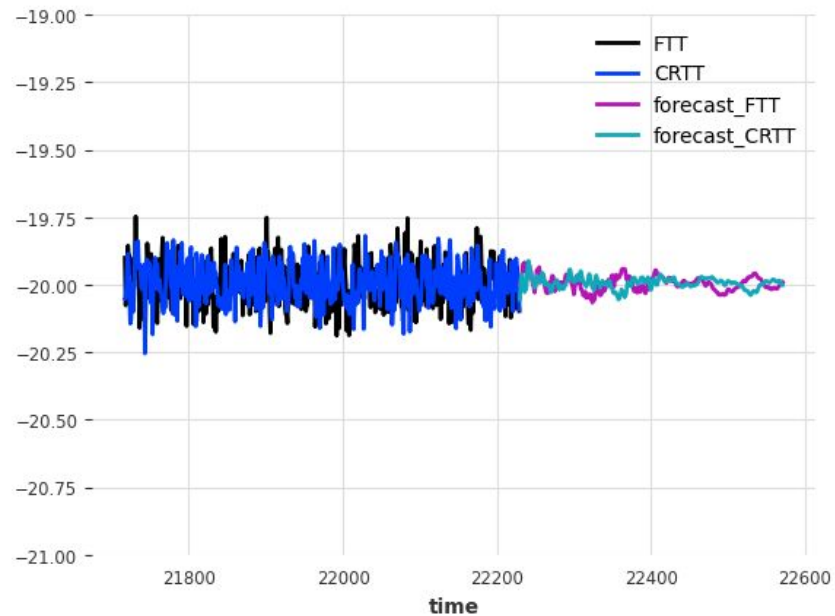


RandomForest

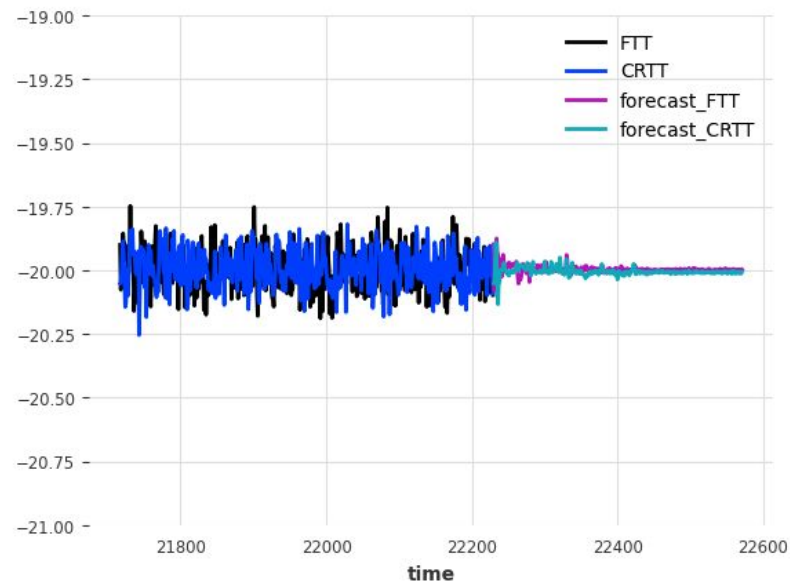


Прогнозирование: отличия

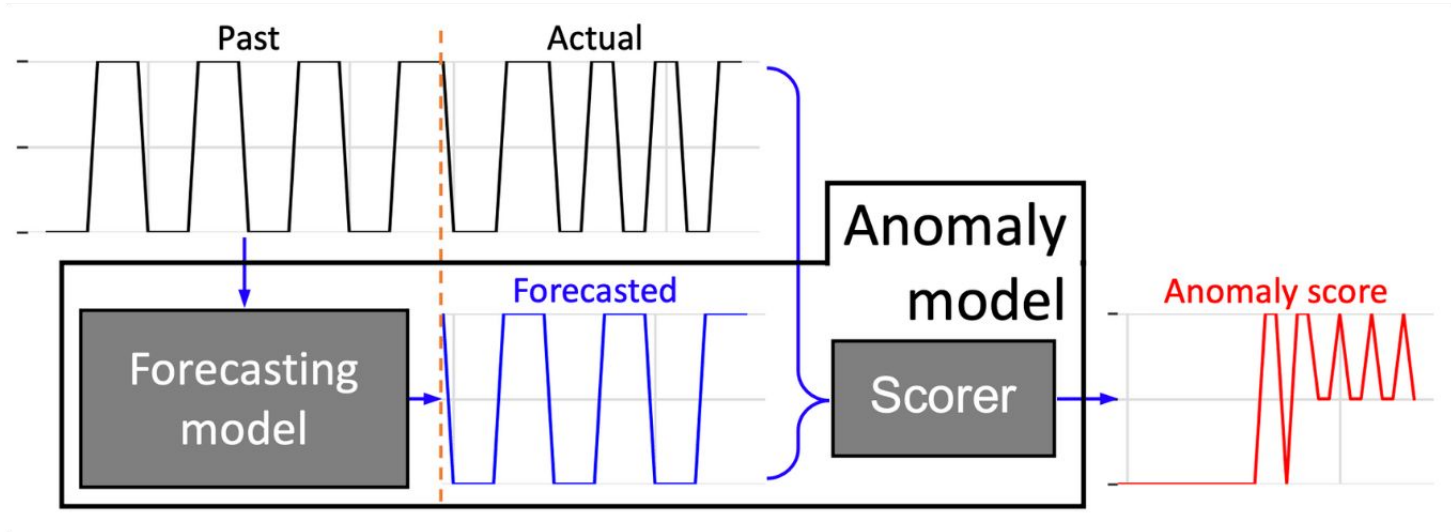
LinearRegression



RandomForest



Обнаружение аномалий



Обнаружение аномалий

Предсказанный временной ряд сравнивается с настоящим. Для количественного описания были выбраны два Scorer'a.

1. NormScorer

Возвращает поэлементную норму некоторого порядка (первого, например) между точками данных прогнозируемого и фактического временных рядов.

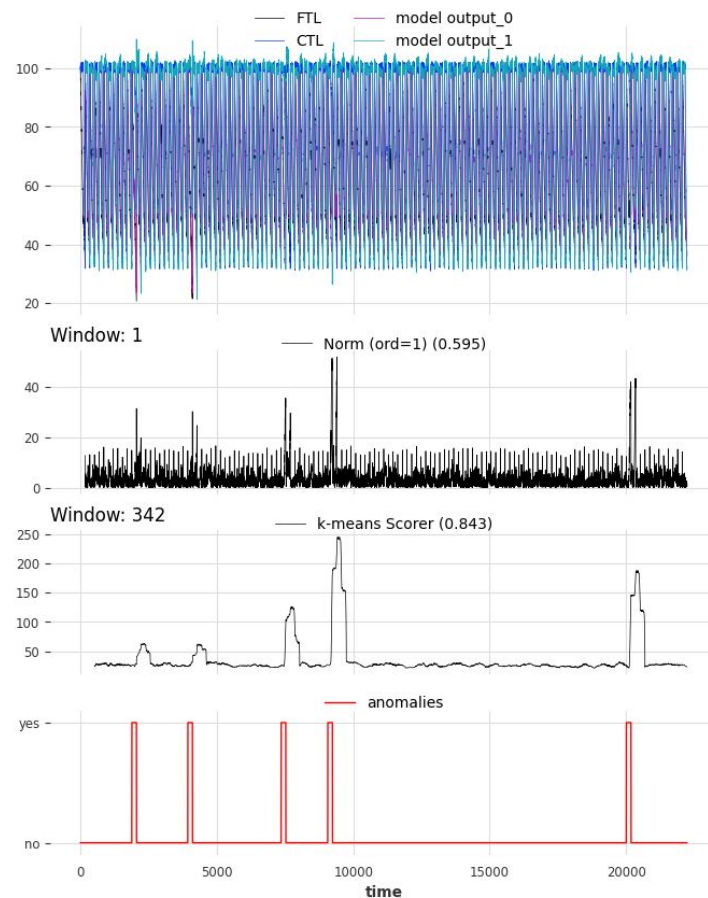
2. KMeansScorer

- a. Применяет скользящее окно к разнице между временными рядами.
- b. Используя метод k-средних, выделяет кластеры на основе полученных векторов.
- c. Для подсчета очков используется одно и то же движущееся окно и возвращает расстояние до ближайшего из k центров кластеров для каждого вектора.
- d. На дополнительном этапе постобработки эти присвоенные баллы преобразуются в баллы аномалий по точкам путем вычисления среднего значения баллов перекрывающихся окон.

Linear Regression (level)

- MAE: 1.68
- RMSE: 2.68

Все аномалии распознались, можно выделить типы аномалий.

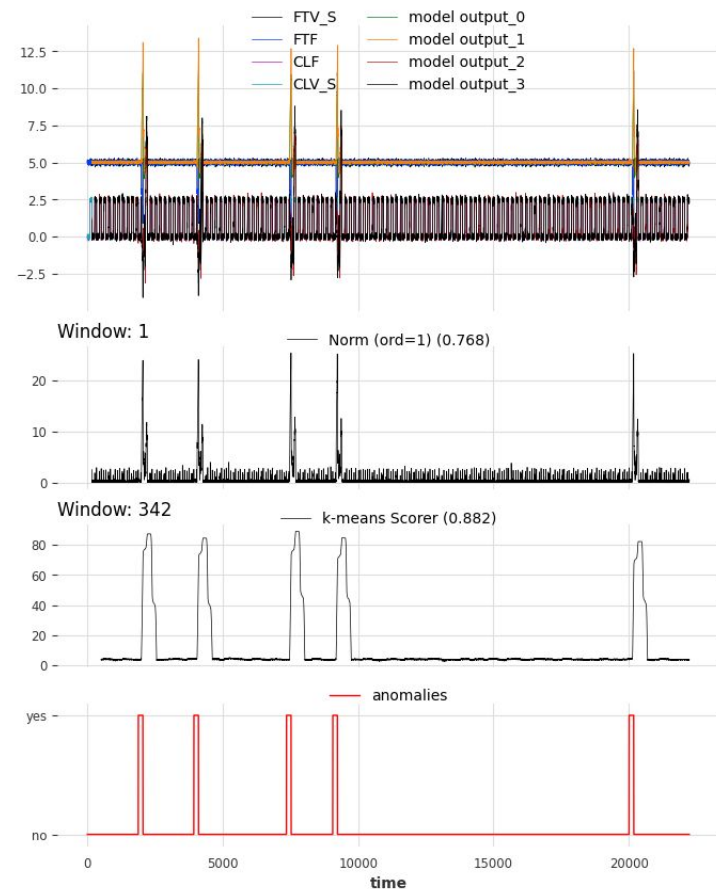


Linear Regression (flow)

- MAE: 0.176
- RMSE: 0.663

Все аномалии распознались, однако
выделить типы аномалий проблематично.

Anomaly results (LinearRegressionModel)

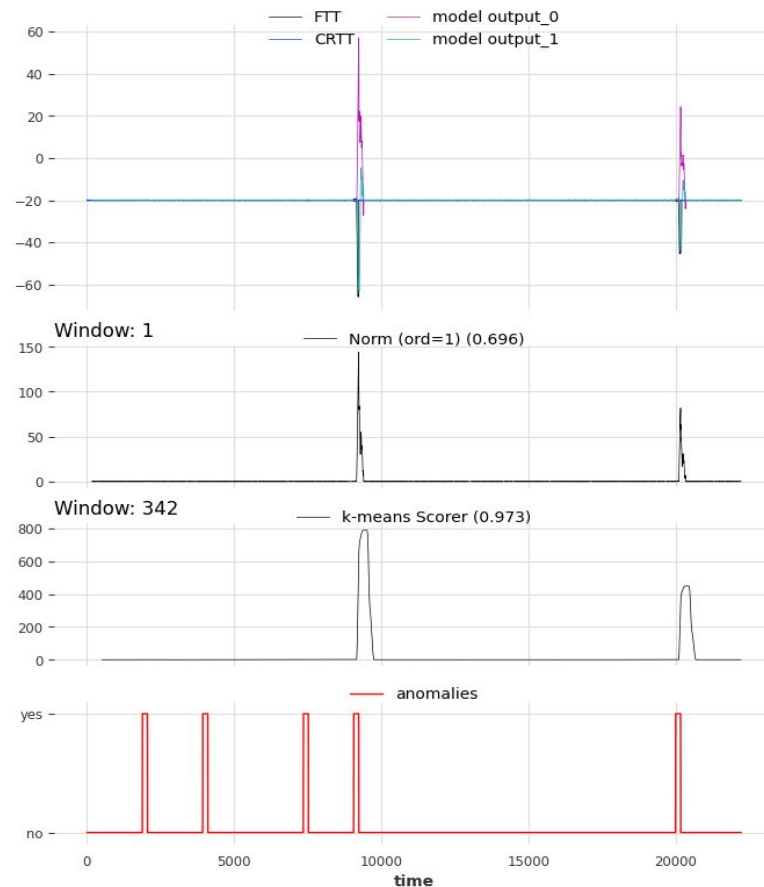


Linear Regression (temperature)

- MAE: 0.499
- RMSE: 3.98

Распознались не все аномалии.

Anomaly results (LinearRegressionModel)

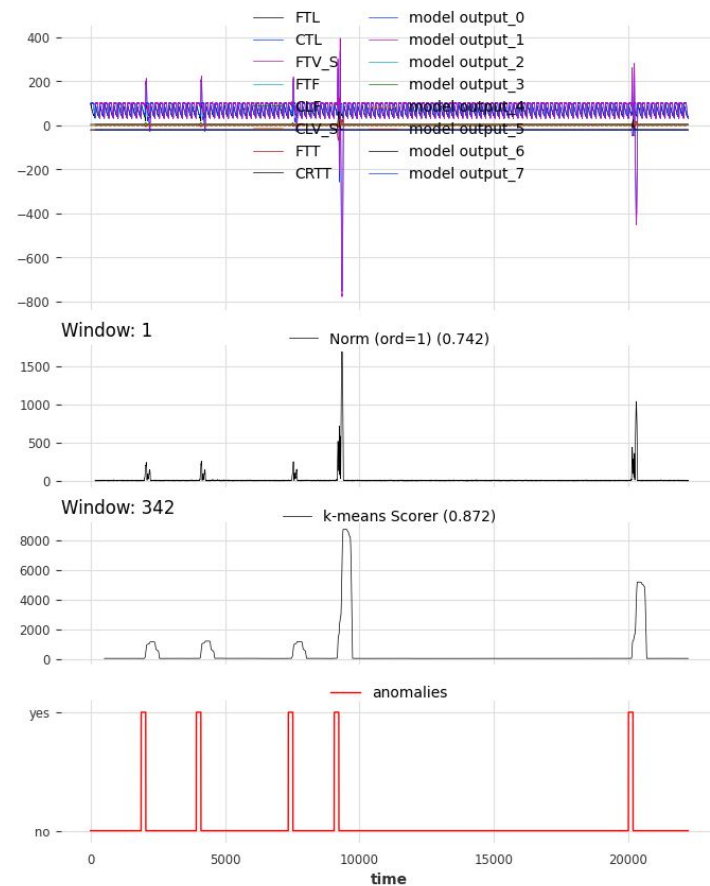


Linear Regression (all params)

- MAE: 2.15
- RMSE: 13.7

Самые большие ошибки, однако линейная модель учитывает данные разных размерностей (% , °C, литры/сек).

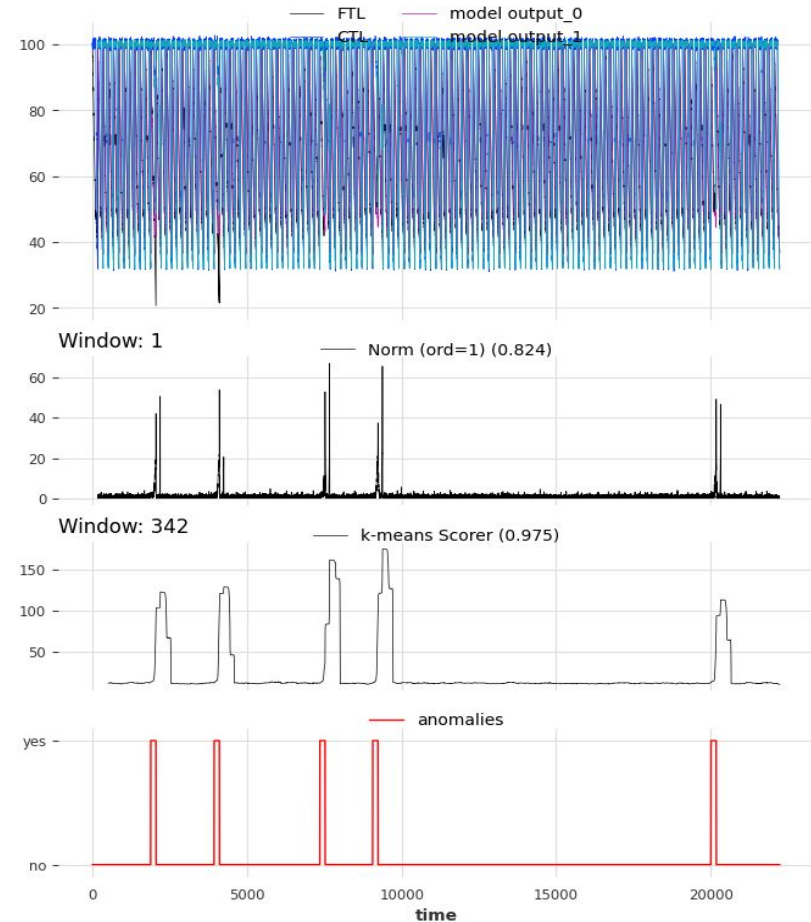
Anomaly results (LinearRegressionModel)



Random Forest (level)

- MAE = 0.566
- MSE = 1.61

Еще более характерны аномалии и их типы.

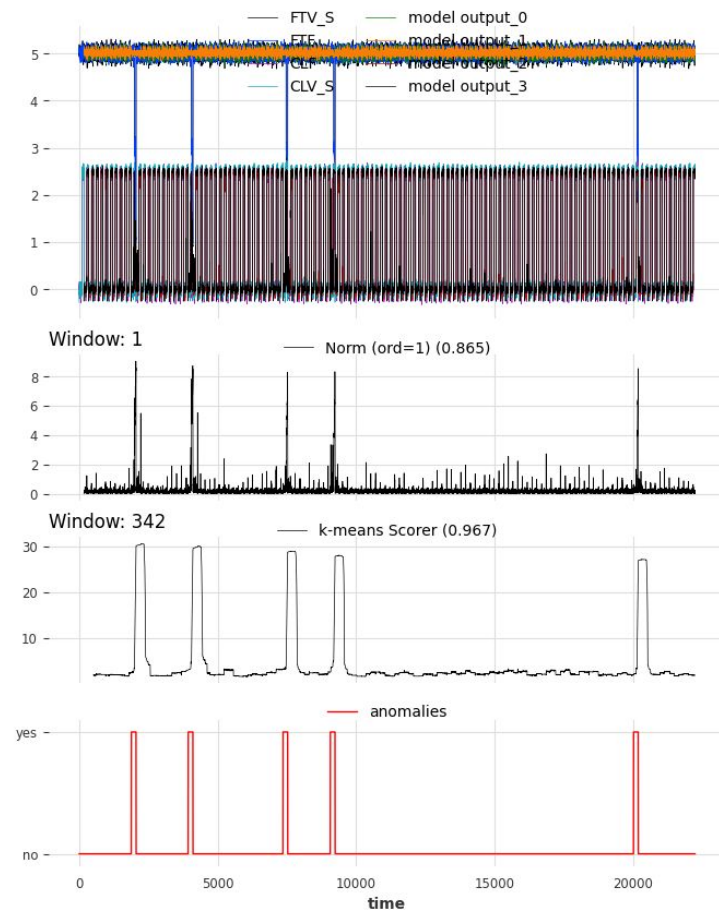


Random Forest (flow)

- MAE = 0.0629
- RMSE = 0.211

Все аномалии распознались, однако выделить типы аномалий проблематично.

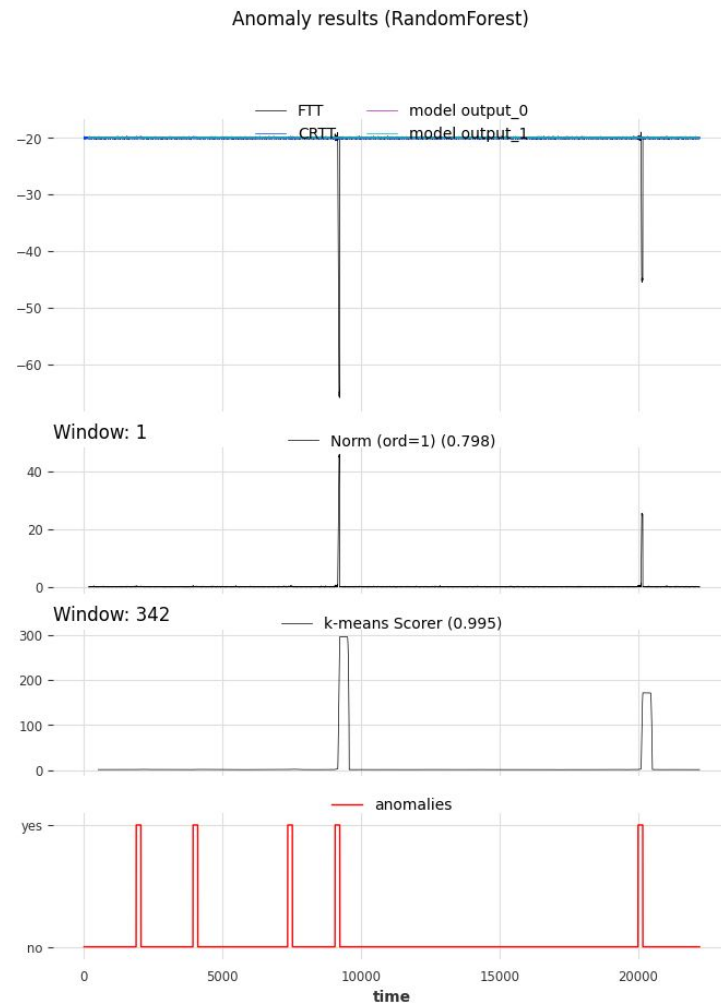
Anomaly results (RandomForest)



Random Forest (temperature)

- MAE = 0.121
- RMSE = 1.18

Распознались не все аномалии.

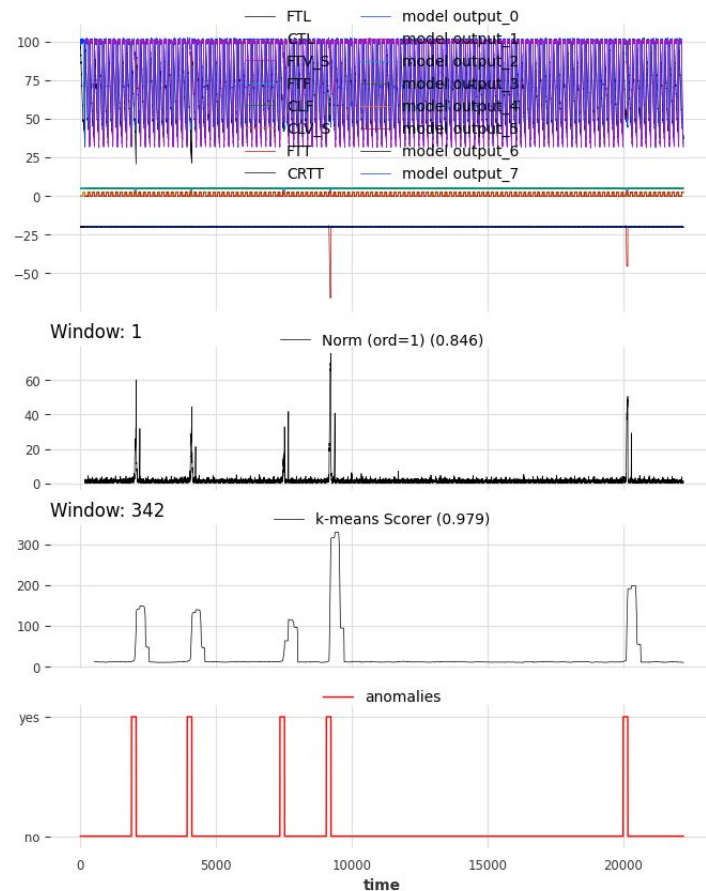


Random Forest (all params)

- MAE = 0.221
- RMSE = 0.796

Для учета данных разных размерностей
RandomForest справился лучше
LinearRegression.

Anomaly results (RandomForest)



Обнаружение аномалий: метрики

| Параметры | LinearRegression (MAE/RMSE) | RandomForest(MAE/RMSE) |
|-------------|-----------------------------|------------------------|
| Level | 1.68/2.68 | 0.566/1.61 |
| Flow | 0.176/0.663 | 0.0629/0.211 |
| Temperature | 0.499/3.98 | 0.121/1.18 |
| All params | 2.15/13.7 | 0.221/0.796 |