SKAB

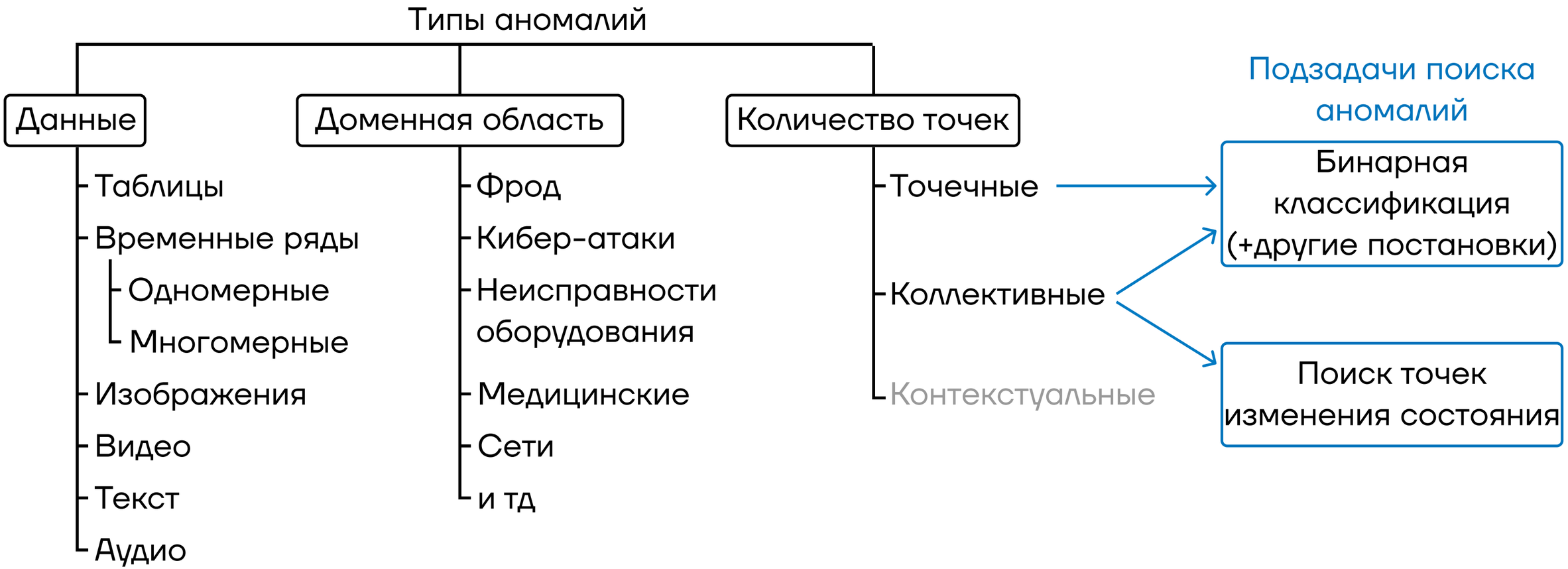
Skoltech
Anomaly
Benchmark

Юрий Кацер, co-founder waico.tech

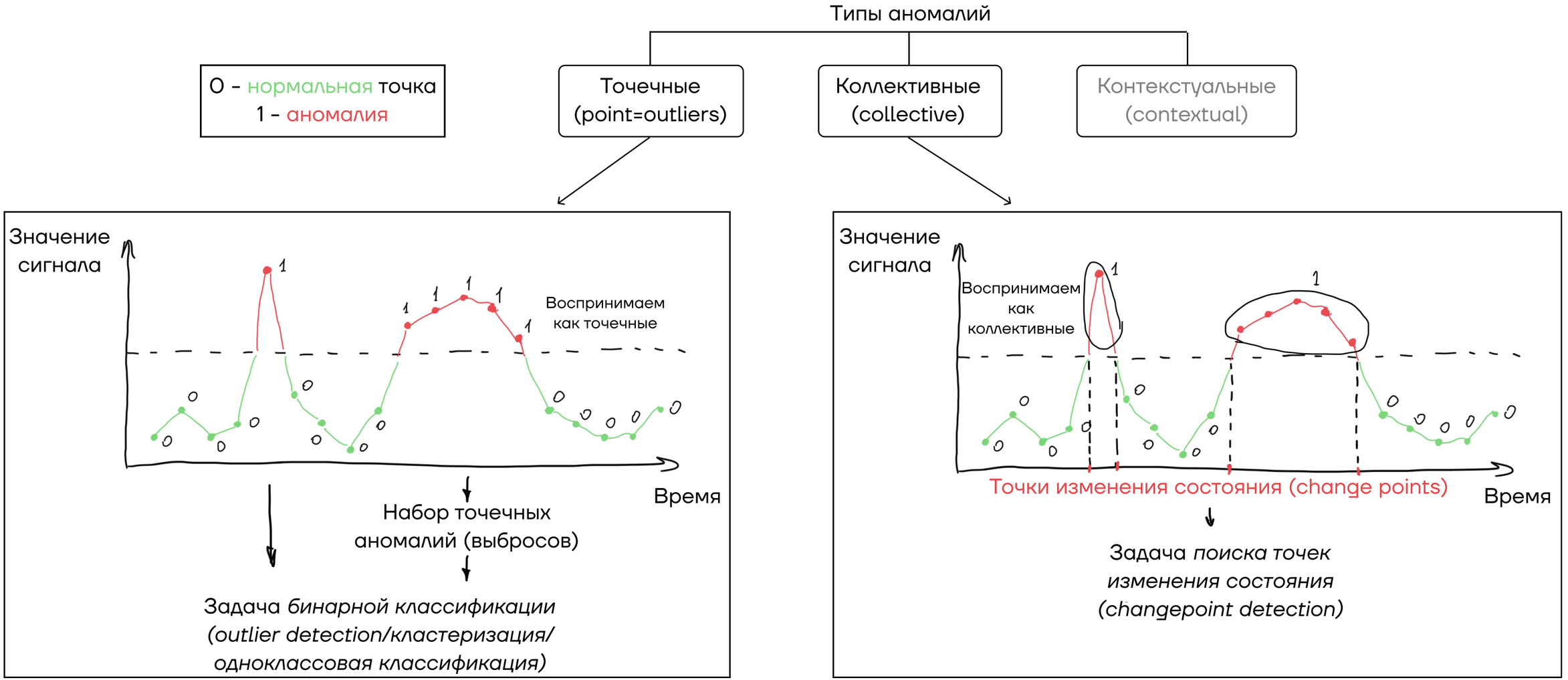
Skoltech

waico.tech

Разнообразие типов аномалий и задач

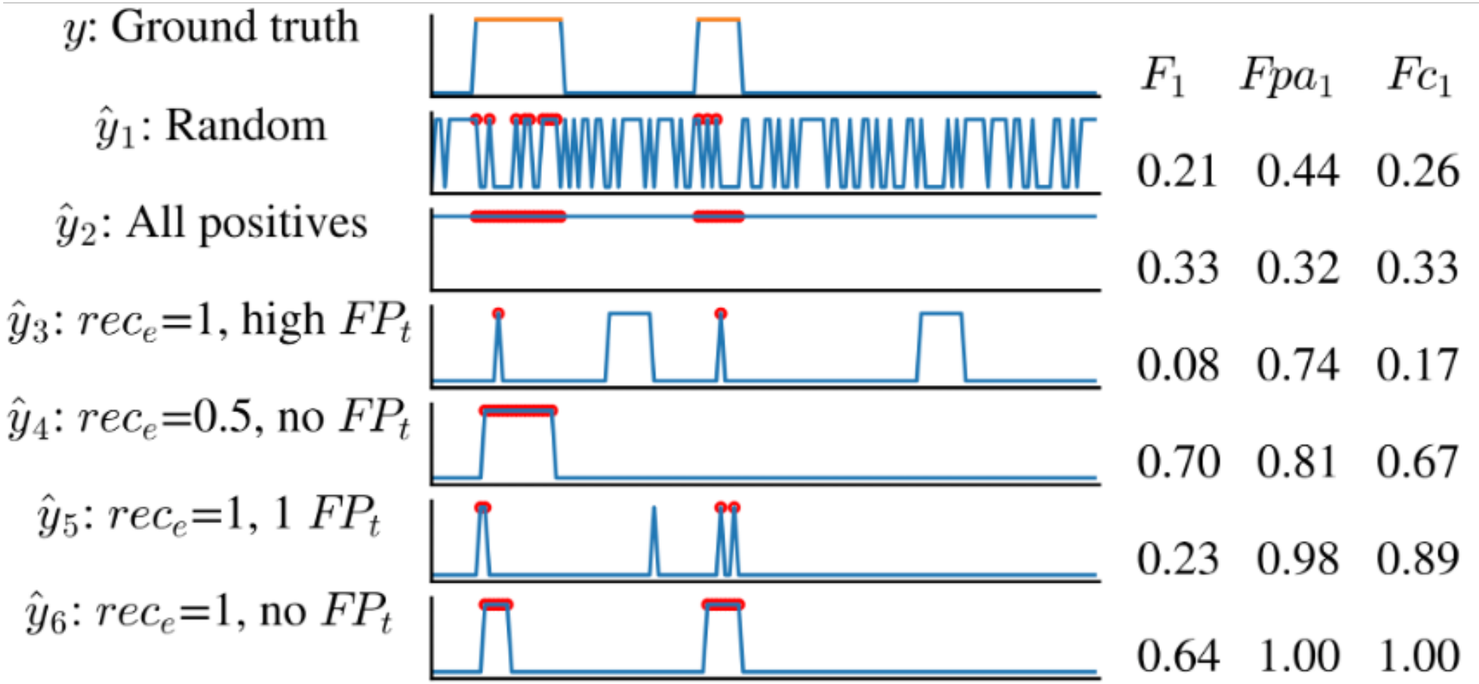


Постановки задач поиска аномалий



Бенчмаркинг

Бенчмарк - датасет, позволяющий исследователям систематически и реалистично оценивать новые алгоритмы по сравнению с существующими альтернативами [Burg2020]



Пример оценки алгоритмов [Garg2021]

van den Burg, Gerrit JJ, and Christopher KI Williams. "An evaluation of change point detection algorithms." arXiv preprint arXiv:2003.06222 (2020).
Garg, Astha, et al. "An Evaluation of Anomaly Detection and Diagnosis in Multivariate Time Series." IEEE Transactions on Neural Networks and Learning Systems (2021).

Актуальность и обзор бенчмарков

-Небольшое количество бенчмарков в области технической диагностики

-Большинство сущствующих бенчмарков направлены на задачу бинарной классификации

-Недостаточность информации (лицензия, задачи, методика оценки, публичный лидерборд и тд)

Dataset	Application	Anomaly type	License	Year
KDD Cup 99	Intrusion, Fraud	Point	N.I.	1999
DARPA 1999	Intrusion, Fraud	Point	N.I.	1999
A New Gas Pipeline dataset	Intrusion	Point, collective	N.I.	2014
S5 (yahoo)	Intrusion, Fraud	Point, collective	Non-commercial research purposes	2015
Microsoft	Intrusion	N.I.	Unpublished	N.I.
KPI (AIOPS data competition)	Web	Collective	N.I.	2018
N-BaIoT: Data for network based detection of IoT botnet attacks	Intrusion	Point	N.I.	2018
Secure Water Treatment (SWaT)	Intrusion	Point, collective	CC BY-NC-SA	2016
Technical benchmarks containing industrial faults and failures				
Delft pump	Faults	Point	N.I.	1999
Tennessee Eastman Process (TEP)	Faults	Collective	N.I.	1993
Sugar Refinery Benchmark (DAMADICS)	Faults	Collective	N.I.	2006
Faults in a urban waste water treatment plant	Faults	Point	N.I.	1993
The Numenta Anomaly Benchmark (NAB)	Intrusion, Fraud, Faults, Web	Collective	GNU AGPL v3.0	2015
Satellite Anomalies	Faults	Point, collective	N.I.	1988
Anomaly Detection in Wireless Sensor Networks	Faults	Collective	Permission is granted to use in any format	2010
Machinery Fault Database	Faults	Collective	N.I.	2016

Актуальность и обзор бенчмарков

-Небольшое число точек изменения состояния

-Одномерные данные

-Эмулированные и симулированные данные

-Разметка для 2х классов (норма, аномалия)

-Очевидные аномалии, обнаруживаемые с помощью “one-liners”

Dataset	Instances	Anomaly instances (CPs)	Duration	Features	Classes (Subclasses)	Kind of data
KDD Cup 99	1,386,021*	512,614*	9 w	41	5(39)**	Emulated
DARPA 1999	—	211 attacks	5 w	8	5(56)**	Emulated
A New Gas Pipeline dataset	274,627	35°	N.I.	6(20)@	5(8)**	Emulated
S5 (yahoo)	367×1,500†	3,915(208)	2 M	1	2	Real-world, Emulated
Microsoft	372×172†	1,871	6 M	1	2	Real-world
KPI (AIOPS data competition)	58×102,119†	134,114(2,453)	N.I.	1	2	Real-world
N-BaIoT: Data for network based detection of IoT botnet attacks	7,062,606	55,624	N.I.	115	11	Real-world
Secure Water Treatment (SWaT)	946,722*	36°	11 d	51*	5	Emulated
Technical benchmarks containing industrial faults and failures						
Delft pump***	1,500	1,124	N.I.	64	2	Emulated
Tennessee Eastman Process (TEP)‡	22×1,460	16,800(21)	3 d	52	22	Emulated
Sugar Refinery Benchmark (DAMADICS)	25×86,400	44,901(37)	1 d	33	20	Emulated
Faults in a urban waste water treatment plant	527	9	527 d	38	13	Real-world
The Numenta Anomaly Benchmark (NAB)	58×6,302†	N.I.(120)	—	1	2	Real-world, Emulated
Satellite Anomalies	5,033	5,032	N.I.	152	9	Real-world
Anomaly Detection in Wireless Sensor Networks Single	4,417	117(2)	6 h	4	2	Real-world
Anomaly Detection in Wireless Sensor Networks Multi	4,690	99(2)	6 h	4	2	Real-world
Machinery Fault Database	1951×250,000	475,5M	5 s	8	2(6)**	Real-world
SKAB (v0.9)	34 × 1,200	13,241(120)	20 m	8	2(7)**	Real-world
SKAB (v1.0)	310 × 1,200	130,200(1240)	20 m	8	2(10)**	Real-world

* - number of unique instances, including test and train sets; ** - number of classes (number of subclasses), e.g. 5 (8) ; *** - accounted dataset ”Dataset Delft pump 5x3” from [54]; † - mean value; @ - 6 features in a raw dataset, 20 features in a preprocessed one [10]; ‡ - a widespread benchmark based on TE process from [17] is accounted; * - reflects only to physical properties of the testbed, there is an additional network traffic data; ° - a total number of single point and collective anomalies.

Цель

Цель исследования - разработать бенчмарк со следующими свойствами:

- Применимый для оценки качества работы алгоритмов в области поиска аномалий технических систем
- Возможность демонстрировать решения обеих подзадач поиска аномалий: поиска выбросов и поиска точек изменения состояния
- Демонстрирующий возможность и значимость результатов оценки на бенчмарке за счет представления бейзлайновых алгоритмов
- Легкая доступность и подробное описание бенчмарка (GitHub, Kaggle и тд)

Испытательный стенд

Стенд состоит из следующий систем:

01 Система циркуляции воды

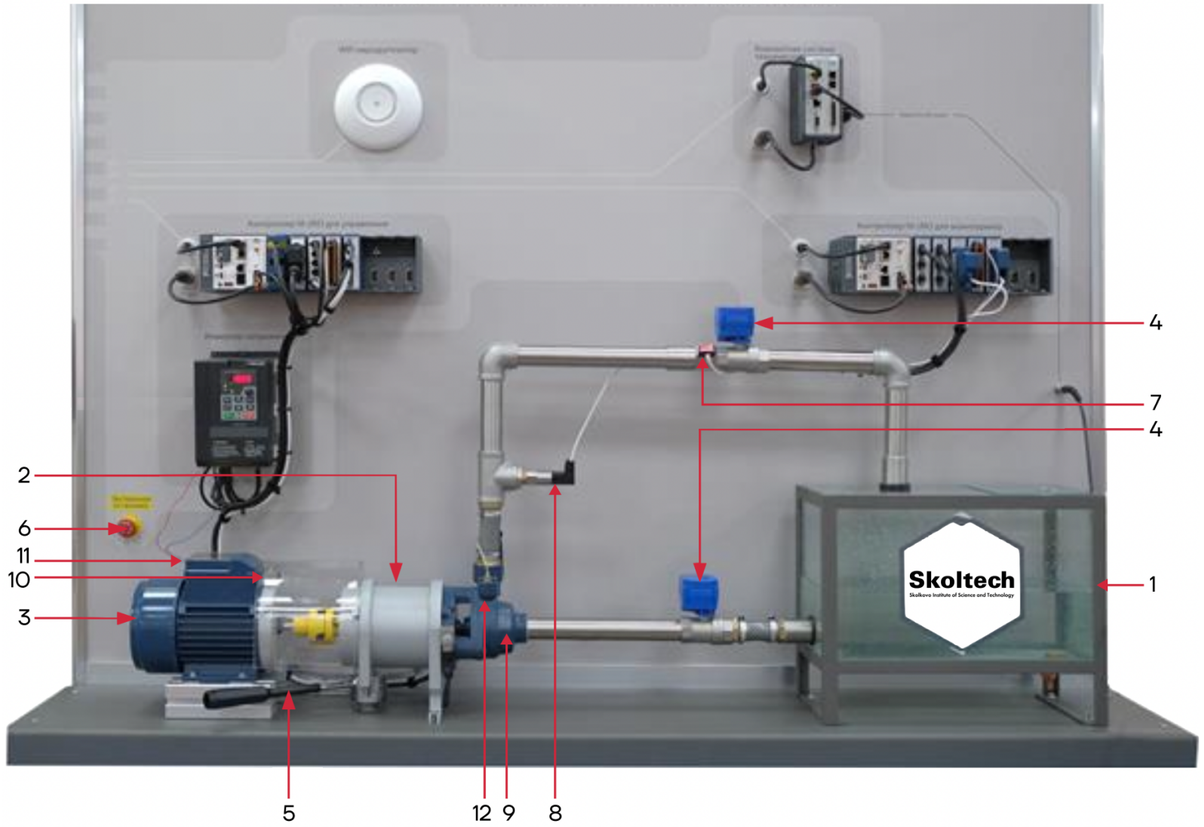
02 Система мониторинга состояния системы циркуляции воды

03 Система контроля и управления системой циркуляции воды

Фокус

04 Система демонстрации технологии Time-Sensitive Networking (TSN)

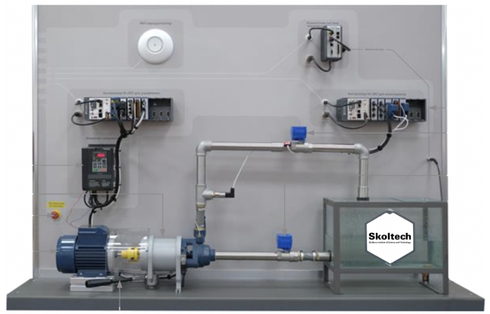
05 Система сбора, обработки и визуализации данных



Передняя панель и состав систем циркуляции воды, мониторинга, контроля и управления:
1 - водяной бак; 2 - водяной насос; 3 - электродвигатель; 4 - клапаны; 5 - механический рычаг для обеспечения несоосности валов; 6 - кнопка аварийной остановки.

Датчики:
7 - датчик расхода (NI 9401 8-channel); 8 - датчик давления (NI 9203 8-channel); 9, 10 - вибродатчики (NI 9232 3-channel); 11, 12 - термопары (NI 9213 Spring Terminal 16-channel thermocouple).

Параметры эксперимента



The Testbed

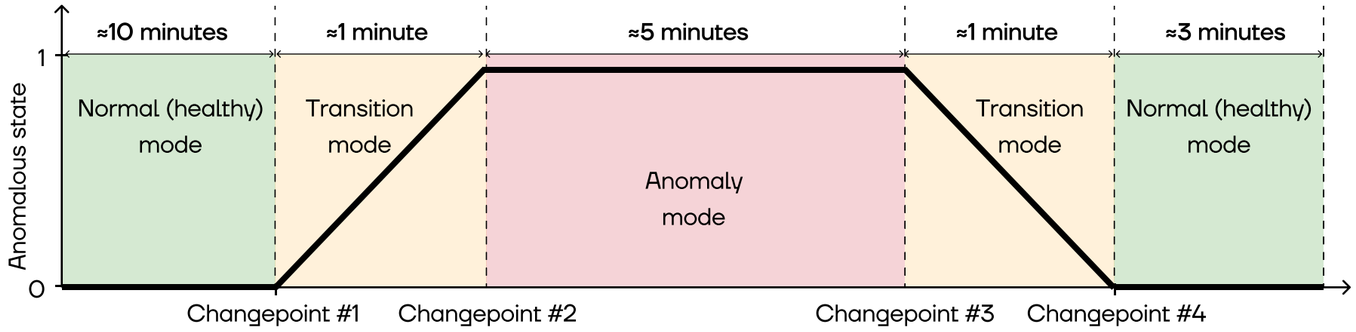
OPC UA

MySQL DB

processing

CSV

.csv files



Каждый .csv файл содержит колонки:

1. datetime - Represents dates and times of the moment when the value is written to the database (YYYY-MM-DD hh:mm:ss)
2. Accelerometer1RMS - Shows an absolute vibration acceleration (Amount of g units)
3. Accelerometer2RMS - Shows an absolute vibration acceleration (Amount of g units)
4. Current - Shows the amperage on the electric motor (Ampere)
5. Pressure - Represents the pressure in the loop after the water pump (Bar)
6. Temperature - Shows the temperature of the engine body (Degree Celsius)
7. Thermocouple - Represents the temperature of the fluid in the circulation loop (Degree Celsius)
8. Voltage - Shows the voltage on the electric motor (Volt)
9. RateRMS - Represents the circulation flow rate of the fluid inside the loop (Liter per minute)

features (X)

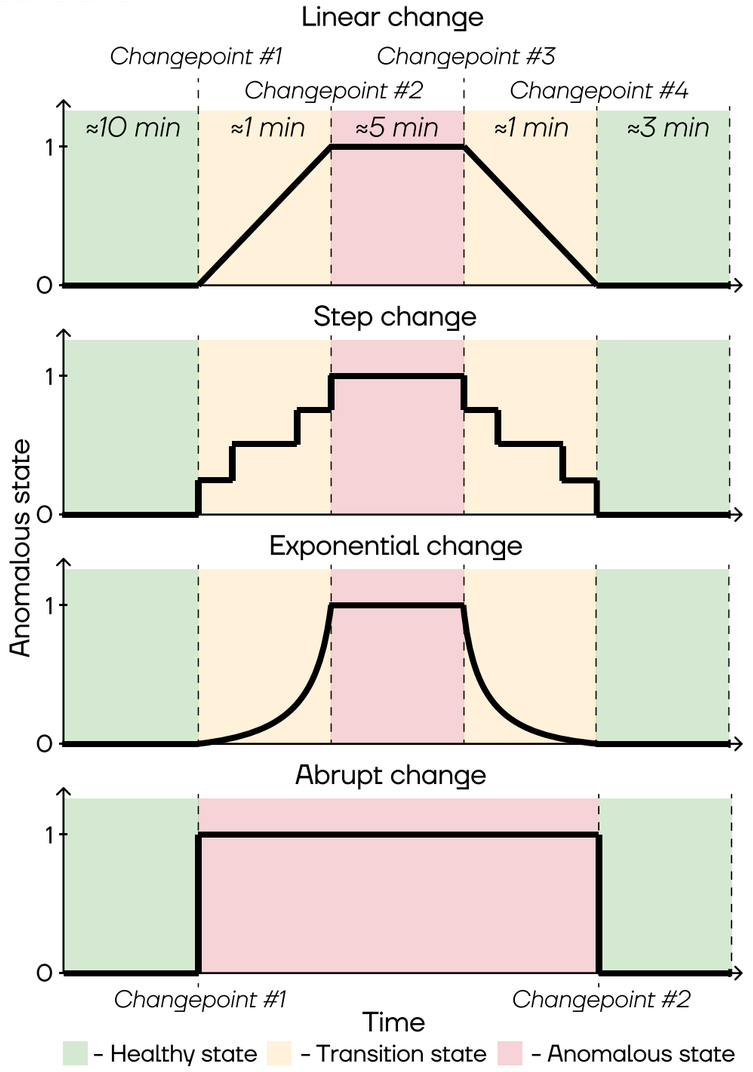
10. anomaly - Shows if the point is anomalous (0 or 1)
11. changepoint - Shows if the point is a changepoint for collective anomalies (0 or 1)

target (y)

Эксперименты и типы аномалий

Anomaly	Total amount	Type
Partly closed valve 1	4† x 4	collective
Partly closed valve 2	4† x 1	collective
Shaft imbalance	5† x 1	collective
Cavitation by means of water-air flow moving to pump	1	collective
Cavitation by means of leakage water from tank	1	collective
Faults associated with adding water to tank	6	collective
Increase and decrease temperature of water in tank	1	collective
Total anomalies	34	

† refers to the existence of different strategies for anomaly appearance obtaining that are shown in Fig. 3 on the example of partly closed valve anomaly. These strategies might make it possible to judge the degree of sensitivity of the algorithms to anomalies.



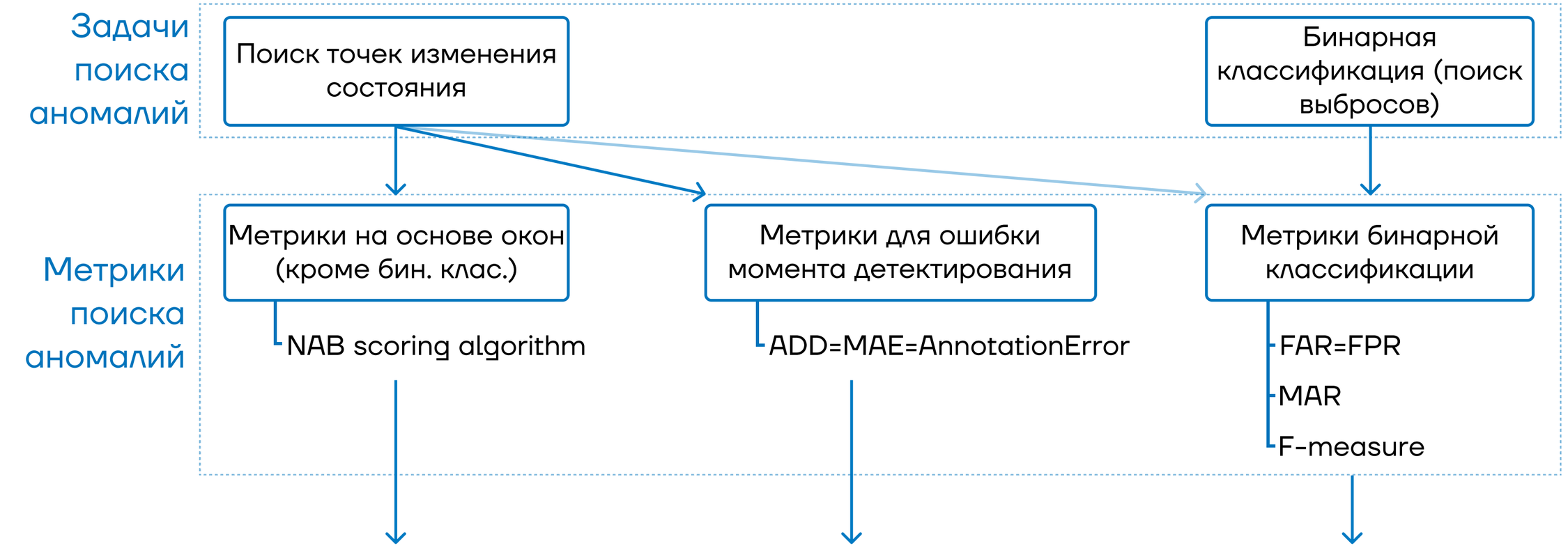
+ Датасет с длительной “нормальной” эксплуатацией

Сравнение с существующими бенчмарками

Dataset	Instances	Anomaly instances (CPs)	Duration	Features	Classes (Subclasses)	Kind of data
KDD Cup 99	1,386,021*	512,614*	9 w	41	5(39)**	Emulated
DARPA 1999	—	211 attacks	5 w	8	5(56)**	Emulated
A New Gas Pipeline dataset	274,627	35◇	N.I.	6(20)@	5(8)**	Emulated
S5 (yahoo)	367×1,500†	3,915(208)	2 M	1	2	Real-world, Emulated
Microsoft	372×172†	1,871	6 M	1	2	Real-world
KPI (AIOPS data competition)	58×102,119†	134,114(2,453)	N.I.	1	2	Real-world
N-BaIoT: Data for network based detection of IoT botnet attacks	7,062,606	55,624	N.I.	115	11	Real-world
Secure Water Treatment (SWaT)	946,722*	36◇	11 d	51*	5	Emulated
Technical benchmarks containing industrial faults and failures						
Delft pump***	1,500	1,124	N.I.	64	2	Emulated
Tennessee Eastman Process (TEP)‡	22×1,460	16,800(21)	3 d	52	22	Emulated
Sugar Refinery Benchmark (DAMADICS)	25×86,400	44,901(37)	1 d	33	20	Emulated
Faults in a urban waste water treatment plant	527	9	527 d	38	13	Real-world
The Numenta Anomaly Benchmark (NAB)	58×6,302†	N.I.(120)	—	1	2	Real-world, Emulated
Satellite Anomalies	5,033	5,032	N.I.	152	9	Real-world
Anomaly Detection in Wireless Sensor Networks Single	4,417	117(2)	6 h	4	2	Real-world
Anomaly Detection in Wireless Sensor Networks Multi	4,690	99(2)	6 h	4	2	Real-world
Machinery Fault Database	1951×250,000	475,5M	5 s	8	2(6)**	Real-world
SKAB (v0.9)	34 × 1,200	13,241(120)	20 m	8	2(7)**	Real-world
SKAB (v1.0)	310 × 1,200	130,200(1240)	20 m	8	2(10)**	Real-world

* - number of unique instances, including test and train sets; ** - number of classes (number of subclasses), e.g. 5 (8) ; *** - accounted dataset "Dataset Delft pump 5x3" from [54]; † - mean value; @ - 6 features in a raw dataset, 20 features in a preprocessed one [10]; ‡ - a widespread benchmark based on TE process from [17] is accounted; * - reflects only to physical properties of the testbed, there is an additional network traffic data; ◇ - a total number of single point and collective anomalies.

Метрики и алгоритмы оценки




```
# nab metric calculation
nab = evaluating_change_point(true_cp,
                              predicted_cp,
                              metric='nab',
                              numenta_time='30 sec')
```

```
Standart  -  17.87
LowFP  -  3.44
LowFN  -  23.2
```

```
# average detection delay metric calculation
add = evaluating_change_point(true_cp,
                              predicted_cp,
                              metric='average_delay',
                              numenta_time='30 sec')
```

```
Average delay 0 days 00:00:07.150000
A number of missed CPs = 90
```

```
# binary classification metrics calculation
binary = evaluating_change_point(true_outlier,
                                 predicted_outlier,
                                 metric='binary',
                                 numenta_time='30 sec')
```

```
False Alarm Rate 12.14 %
Missing Alarm Rate 52.56 %
F1 metric 0.56
```

Алгоритмы для начального лидерборда

Алгоритмы поиска выбросов и точек изменения состояния:

1. **T2.** Statistical process control charts are quite common techniques for process monitoring. T2 Hotteling's statistics is a multivariate control chart based on the mahalanobis distance between current state vector and the target vector which represents normal state of a process.
2. **T2 +Q (PCA-based).** Principal component analysis (PCA) is one of the most popular techniques for process monitoring and anomaly detection. PCA is often used as a part of anomaly detection pipeline based on T2 and Q indicators generation and analysis. In this case T2 statistics is based on the first m of n principal components, while Q statistics is based on the remaining n-m PCs.
3. **LSTM-based NN (LSTM - Long-short term memory).** We use LSTM-based approach for anomaly detection. The error between real data points and LSTM-based 1 point ahead prediction at a time moment i based on Mean Squared Error (MSE) was used for the anomaly detection.
4. **Feed-Forward Autoencoder.** Autoencoder was initially proposed as dimension reduction technique capable of find nonlinear dependencies between the features, but it can be used for anomaly detection problem solving as well. It is a feed-forward symmetric neural network architecture that reconstructs or replicates the data. The reconstruction error can be used as the anomaly score. We propose MSE as the reconstruction error to make this approach be like an LSTM-based one.
5. **Convolutional Autoencoder (Conv-AE).** A reconstruction convolutional autoencoder model to detect anomalies in timeseries data using reconstruction error as an anomaly score.
6. **LSTM Autoencoder (LSTM-AE).** A reconstruction sequence-to-sequence (LSTM-based) autoencoder model to detect anomalies in timeseries data using reconstruction error as an anomaly score.
7. **LSTM Variational Autoencoder (LSTM-VAE).** A reconstruction LSTM variational autoencoder model to detect anomalies in timeseries data using reconstruction error as an anomaly score.
8. **Variational Autoencoder (VAE).** A reconstruction variational autoencoder (VAE) model to detect anomalies in timeseries data using reconstruction error as an anomaly score. VAE is an autoencoder that learns a latent variable model for its input data.
9. **MSCRED.** Subsequently, given the signature matrices, a convolutional encoder is employed to encode the inter-sensor (time series) correlations patterns and an attention based Convolutional Long-Short Term Memory (ConvLSTM) network is developed to capture the temporal patterns.
10. **MSET.** Multivariate state estimation technique is a non-parametric and statistical modeling method, which calculates the estimated values based on the weighted average of historical data.
11. **Isolation forest.** It is a machine learning algorithm that isolates anomalous points from the normal instances using the assumption that "anomalies are those instances which have short average path lengths on the iTrees". This method is relatively simple, intuitive, and it has just a few hyperparameters for the tunning.

Дополнительно:

1. **Perfect detector.** It denotes the algorithm that finds all the anomalies when they occur, and it never fails.
2. **Null detector.** The null detector detects no anomalies at all, showing a normal state for all instances.

Результаты (лидерборды)

Точечные аномалии (бинарная классификация)

Отсортировано по F1; для F1 больше - лучше; и для FAR (False Alarm Rate), и для MAR (Missing Alarm Rate) меньше - лучше

Algorithm	F1	FAR, %	MAR, %
Perfect detector	1	0	0
Conv-AE	***0.79***	13.69	***17.77***
MSET	0.73	20.82	20.08
LSTM-AE	0.68	14.24	35.56
T-squared+Q (PCA)	0.67	13.95	36.32
LSTM	0.64	15.4	39.93
MSCRED	0.64	13.56	41.16
LSTM-VAE	0.56	9.13	55.03
T-squared	0.56	12.14	52.56
Autoencoder	0.45	7.56	66.57
Isolation forest	0.4	***6.86***	72.09
Null detector	0	0	100

Коллективные аномалии (поиск точек изменения состояния)

Отсортировано по NAB (standard); для всех метрик больше - лучше

Algorithm	NAB (standard)	NAB (lowFP)	NAB (LowFN)
Perfect detector	100	100	100
Isolation forest	***37.53***	17.09	***45.02***
MSCRED	28.74	***23.43***	31.21
LSTM	27.09	11.06	32.68
T-squared+Q (PCA)	26.71	22.42	28.32
ruptures**	24.1	21.69	25.04
CPDE***	23.07	20.52	24.35
LSTM-AE	22.12	20.01	23.21
LSTM-VAE	19.17	15.39	20.98
T-squared	17.87	3.44	23.2
ArimaFD	16.06	14.03	17.12
Autoencoder	15.59	0.78	20.91
MSET	12.71	11.04	13.6
Conv-AE	10.09	8.62	10.83
Null detector	0	0	0

Результаты (лидерборды)

Точечные аномалии (бинарная классификация)

Отсортировано по F1; для F1 больше - лучше; и для FAR (False Alarm Rate), и для MAR (Missing Alarm Rate) меньше - лучше

Algorithm	F1	FAR, %	MAR, %
Perfect detector	1	0	0
Conv-AE	***0.79***	13.69	***17.77***
MSET	0.73	20.82	20.08
LSTM-AE	0.68	14.24	35.56
T-squared+Q (PCA)	0.67	13.95	36.32
LSTM	0.64	15.4	39.93
MSCRED	0.64	13.56	41.16
LSTM-VAE	0.56	9.13	55.03
T-squared	0.56	12.14	52.56
Autoencoder	0.45	7.56	66.57
Isolation forest	0.4	***6.86***	72.09
Null detector	0	0	100

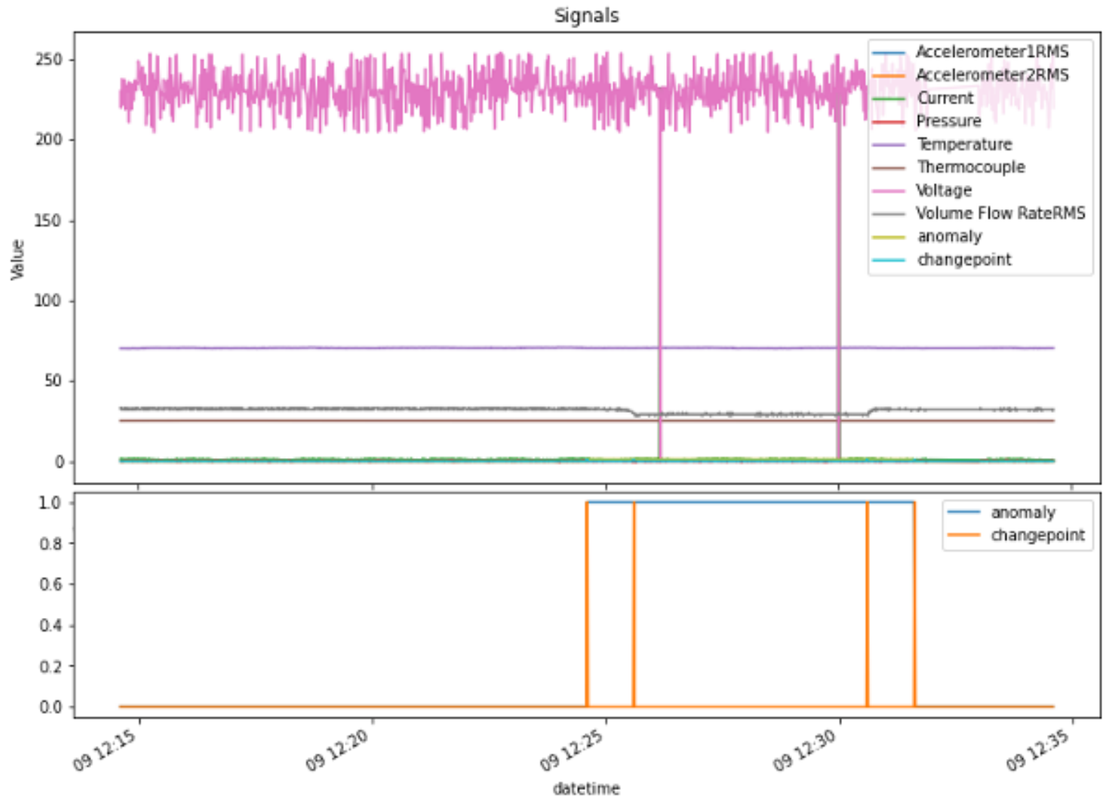
Коллективные аномалии (поиск точек изменения состояния)

Отсортировано по NAB (standard); для всех метрик больше - лучше

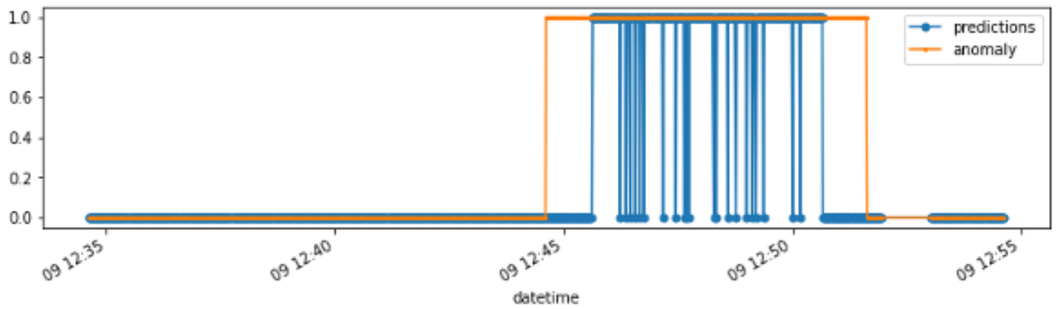
Algorithm	NAB (standard)	NAB (lowFP)	NAB (LowFN)
Perfect detector	100	100	100
Isolation forest	***37.53***	17.09	***45.02***
MSCRED	28.74	***23.43***	31.21
LSTM	27.09	11.06	32.68
T-squared+Q (PCA)	26.71	22.42	28.32
ruptures**	24.1	21.69	25.04
CPDE***	23.07	20.52	24.35
LSTM-AE	22.12	20.01	23.21
LSTM-VAE	19.17	15.39	20.98
T-squared	17.87	3.44	23.2
ArimaFD	16.06	14.03	17.12
Autoencoder	15.59	0.78	20.91
MSET	12.71	11.04	13.6
Conv-AE	10.09	8.62	10.83
Null detector	0	0	0

Пример (T^2 статистика)

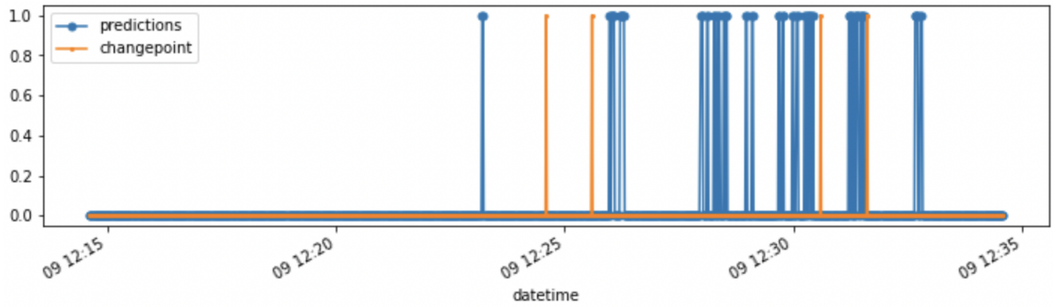
Исходные данные (dataset #1)



Данные на выходе для *бинарной классификации* (dataset #1)



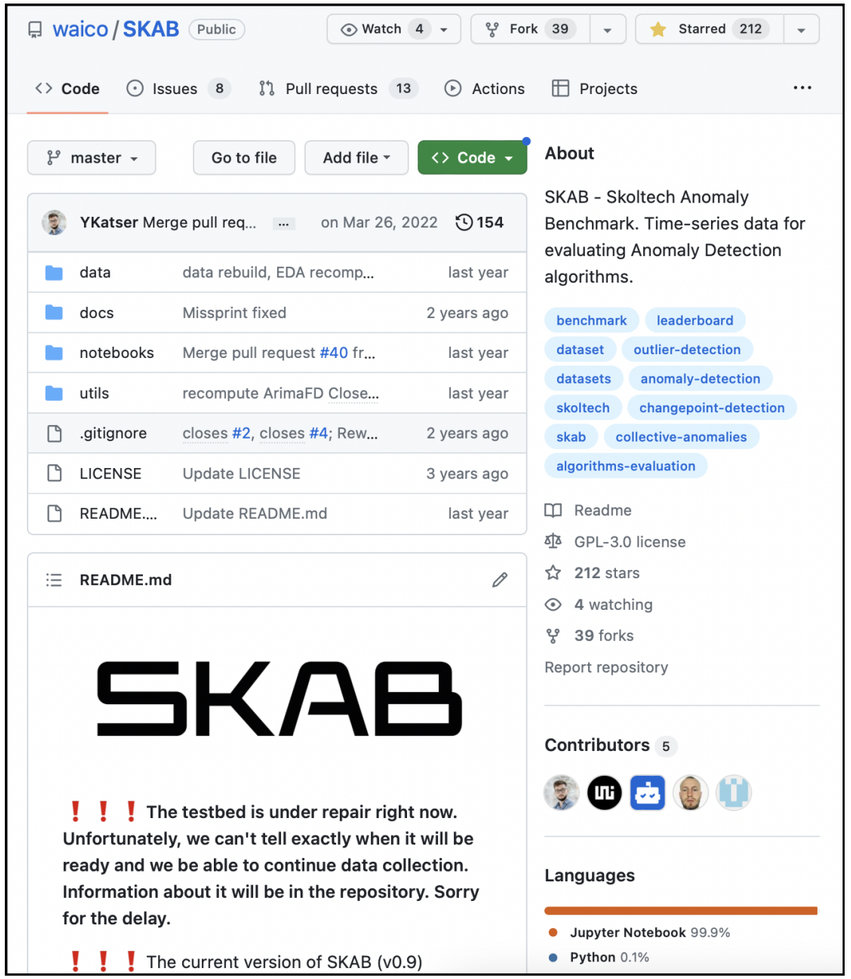
Данные на выходе для *поиска точек изменения состояния* (dataset #1)



Outcomes

1. Датасеты
2. EDA
3. Лидерборды
4. Python модули для оценки результатов
5. Бейзлайны: python блокноты с реализованными алгоритмами обнаружения аномалий
6. 30 цитирований в научных статьях
7. Github* (212 stars) и Kaggle** (37 upvotes) репозитории

1. Статья “SKAB: Skoltech Anomaly Benchmark for Industrial Faults and Failures Detection”
2. Инструкция-статья на medium



Skoltech

waico.tech

* https://github.com/waico/SKAB/
** Iurii D. Katser and Vyacheslav O. Kozitsin, “Skoltech Anomaly Benchmark (SKAB).” Kaggle, 2020, doi: 10.34740/KAGGLE/DSV/1693952.

Спасибо за внимание!

Юрий Кацер, co-founder waico.tech

tg channel @DataKatser