



Московский институт электроники и
математики им. А.Н. Тихонова

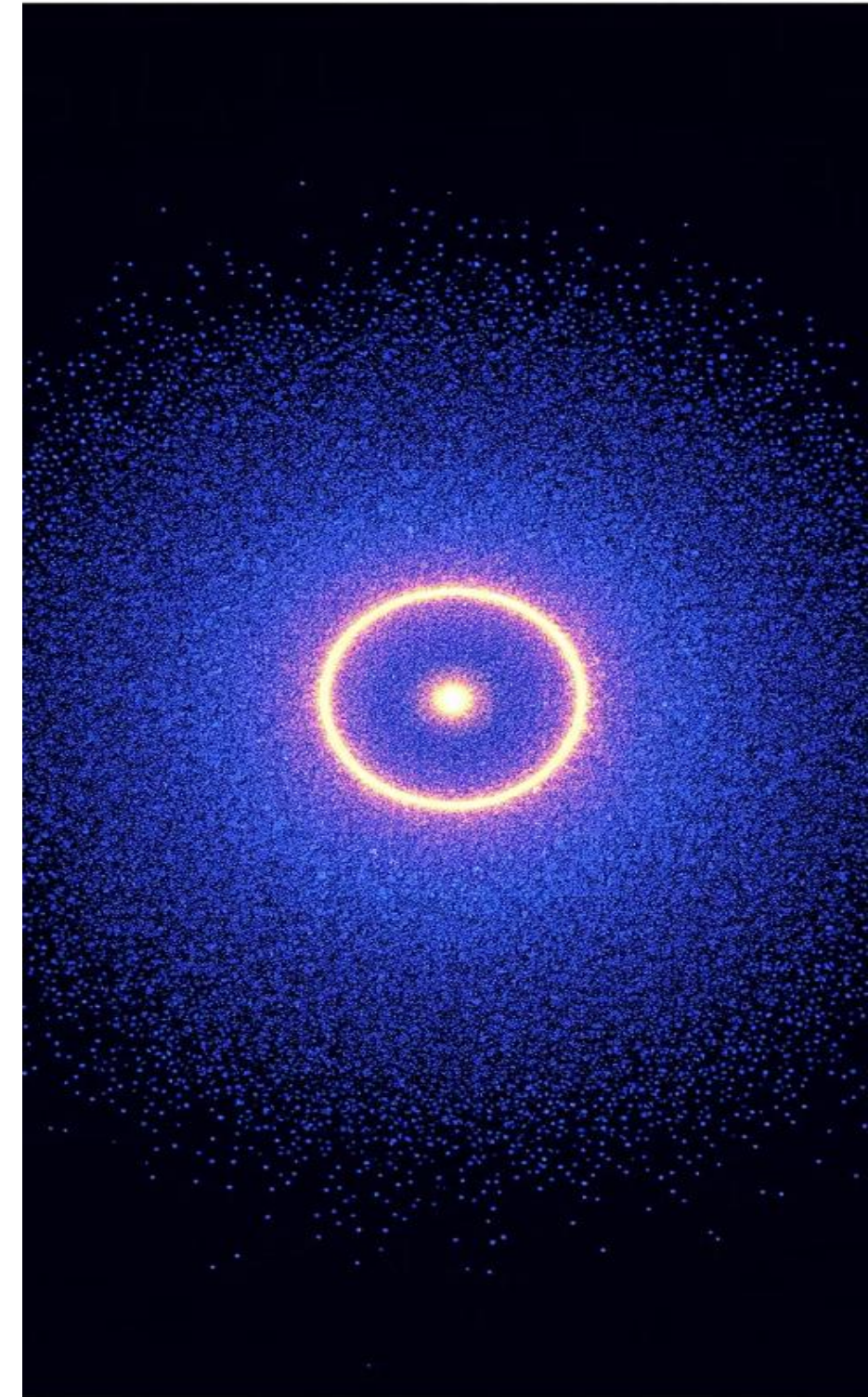
Кафедра информационной
безопасности киберфизических
систем

Москва 2025

Анализ и обнаружение аномалий в IoT-данных

Понятие аномалий в данных

Аномалия – это наблюдение, отклоняющееся от общих закономерностей данных, существенно отличающееся от типичных значений и требующее специального внимания при анализе.



Причины возникновения аномалий



Технические сбои
оборудования



Ошибки датчиков



Внешние факторы



Взлом систем мониторинга



Необычное поведение
наблюдаемого объекта



Задачи анализа аномалий



Предотвращение аварий



Своевременное устранение сбоев



Улучшение качества прогнозов



**Повышение безопасности
и надёжности технологических
процессов и систем**



Типы аномалий в данных



Точечные аномалии

Единичные экстремумы в данных

Контекстуальные аномалии

Значения, аномальные только в определенных условиях

Коллективные аномалии

Группы нетипичных значений в данных



Методы анализа аномалий



Статистические подходы

- Z-score
- Межквартильный размах

Методы машинного обучения

- Isolation Forest
- LOF (Local Outlier Factor)

Визуальный анализ

Графики, диаграммы и другие методы визуализации данных



Isolation Forest: основной принцип

Построение случайных деревьев решений

Алгоритм создает множество деревьев решений со случайным выбором признаков

Определение глубины изоляции

Измеряется, насколько быстро объект изолируется в дереве

Сравнение с нормальными объектами

Аномалии изолируются на меньшей глубине, чем нормальные объекты.

Особенности IoT-данных



Большой объём

Огромные массивы данных от множества устройств



Разнообразие сенсоров

Различные типы датчиков и измеряемых параметров



Высокая частота измерений

Постоянный поток данных в реальном времени



Периодические тренды

Циклические паттерны в данных



Множественные корреляции

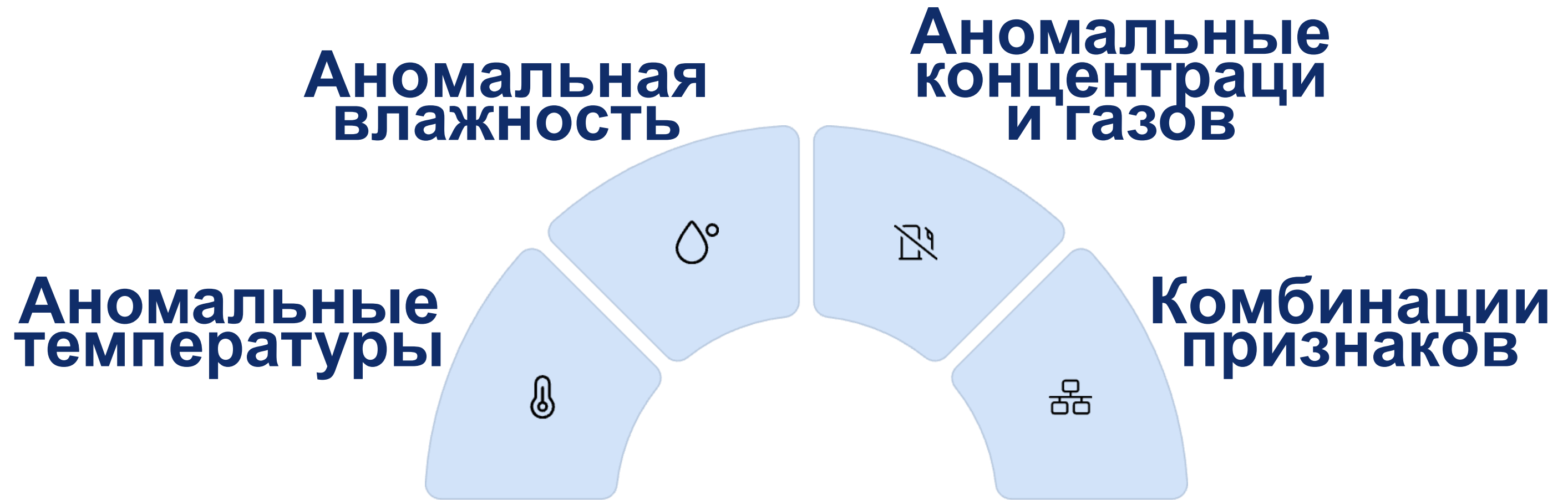
Взаимосвязи между различными параметрами



Наличие шума

Помехи и искажения в измерениях

Задача выявления аномалий в IoT-данных



Возможности RapidMiner для анализа IoT-данных



Загрузка и предобработка данных



Расширения для обнаружения аномалий



Визуализация результатов



Интерпретация результатов

Этапы предобработки IoT-данных



Проверка типов данных

Убедиться, что все данные имеют корректный формат



Обработка пропусков

Заполнение или удаление отсутствующих значений



Преобразование временных меток

Конвертация Epoch в читаемый формат даты и времени



Исключение нерелевантных признаков

Удаление параметров, не влияющих на анализ





Загрузка данных в RapidMiner

Оператор Read CSV

Используется для загрузки
IoT-данных в формате CSV

- Автоматически определяет
типы данных
- Позволяет настроить
параметры импорта

Row No.	ts	device	co	humidity	light	lpg	motion	smoke	temp
1	159451209...	b8:27:eb:bf...	0.005	51	false	0.008	false	0.020	22.700
2	159451209...	00:0f:00:70...	0.003	76	false	0.005	false	0.013	19.700
3	159451209...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
4	159451209...	1c:bf:ce:15:...	0.004	76.800	true	0.007	false	0.019	27
5	159451210...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
6	159451210...	1c:bf:ce:15:...	0.004	77.900	true	0.007	false	0.019	27
7	159451210...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
8	159451210...	00:0f:00:70...	0.003	76	false	0.005	false	0.014	19.700
9	159451210...	1c:bf:ce:15:...	0.004	77.900	true	0.007	false	0.018	27
10	159451210...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
11	159451211...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
12	159451211...	1c:bf:ce:15:...	0.004	78	true	0.007	false	0.019	27
13	159451211...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
14	159451211...	1c:bf:ce:15:...	0.004	78	true	0.007	false	0.019	27
15	159451212...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
16	159451212...	00:0f:00:70...	0.003	75.800	false	0.005	false	0.014	19.700
17	159451212...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600
18	159451212...	b8:27:eb:bf...	0.005	50.900	false	0.008	false	0.020	22.600

ExampleSet (405,184 examples,0 special attributes,9 regular attributes)

	Name	Type	Missing	Statistics			Filter (9 / 9 attributes):	Search for Attribute	
Data	ts	Real	0	Min 1594512094.386	Max 1595203417.264	Average 1594858017.297			
Statistics	device	Nominal	0	Least 1c:bf:ce [...] (105918)	Most b8:27:eb [...] (187451)	Values b8:27:eb:bf:9d:51 (187451), 00:0f:00:70:91:0a (111815), ...[1 mo			
Visualizations	co	Real	0	Min 0.001	Max 0.014	Average 0.005			
Annotations	humidity	Real	0	Min 1.100	Max 99.900	Average 60.512			
	light	Nominal	0	Least true (112527)	Most false (292657)	Values false (292657), true (112527)			
	lpg	Real	0	Min 0.003	Max 0.017	Average 0.007			
	motion	Nominal	0	Least true (482)	Most false (404702)	Values false (404702), true (482)			
	smoke	Real	0	Min 0.007	Max 0.047	Average 0.019			
	temp	Real	0	Min 0	Max 30.600	Average 22.454			

Преобразование временных данных



Исходный формат Epoch
Время в секундах с 1970 года



Generate Attributes
Создание новых атрибутов даты/времени



Numerical to Date
Преобразование числа в дату

Generate Attributes

function descriptions

Edit Parameter List: function descriptions

Edit Parameter List: **function descriptions**
List of functions to generate.

column name	function expressions
ts_millis	ts * 1000

Add Entry Remove Entry Apply Cancel

Process Parameters

Numerical to Date

attribute name ts_millis

☐ keep old attribute

time offset 0





Isolation Forest в RapidMiner

Число деревьев

Определяет количество случайных деревьев для построения модели

Размер листьев

Максимальное количество объектов в листовом узле

Выборка bootstrap

Метод формирования обучающих выборок для деревьев

Эвристика выбора признаков

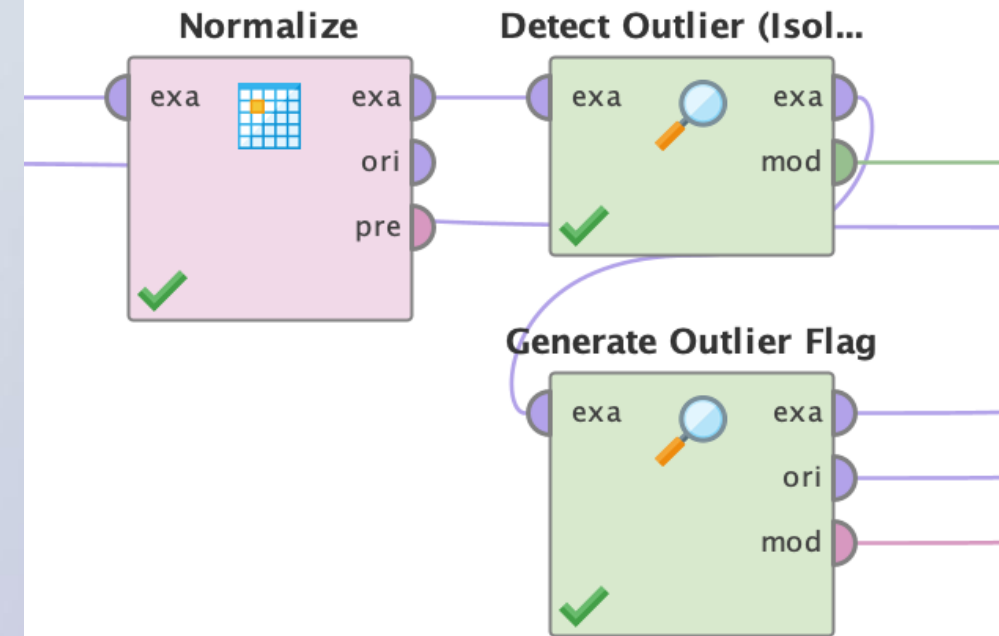
Стратегия отбора признаков при построении деревьев


Process	Parameters
Detect Outlier (Isolation Forest)	
number of trees	100
max leaf size	1
bootstrap ratio	0.9
<input checked="" type="checkbox"/> use feature heuristic	
score calculation	average_path

Оператор Generate Outlier Flag

Функциональность оператора

- Добавляет бинарную метку аномалий
- Выделяет топ-N% наиболее аномальных наблюдений по расчётному скору



Process	Parameters
 Generate Outlier Flag	
<i>method</i>	contamination
<input checked="" type="checkbox"/> <i>define score column</i>	
score column	prediction
contamination threshold	0.05

Визуализация временных аномалий

Линейные графики наглядно отображают аномальные пики в значениях датчиков (например, CO и температуры) с помощью цветовой маркировки

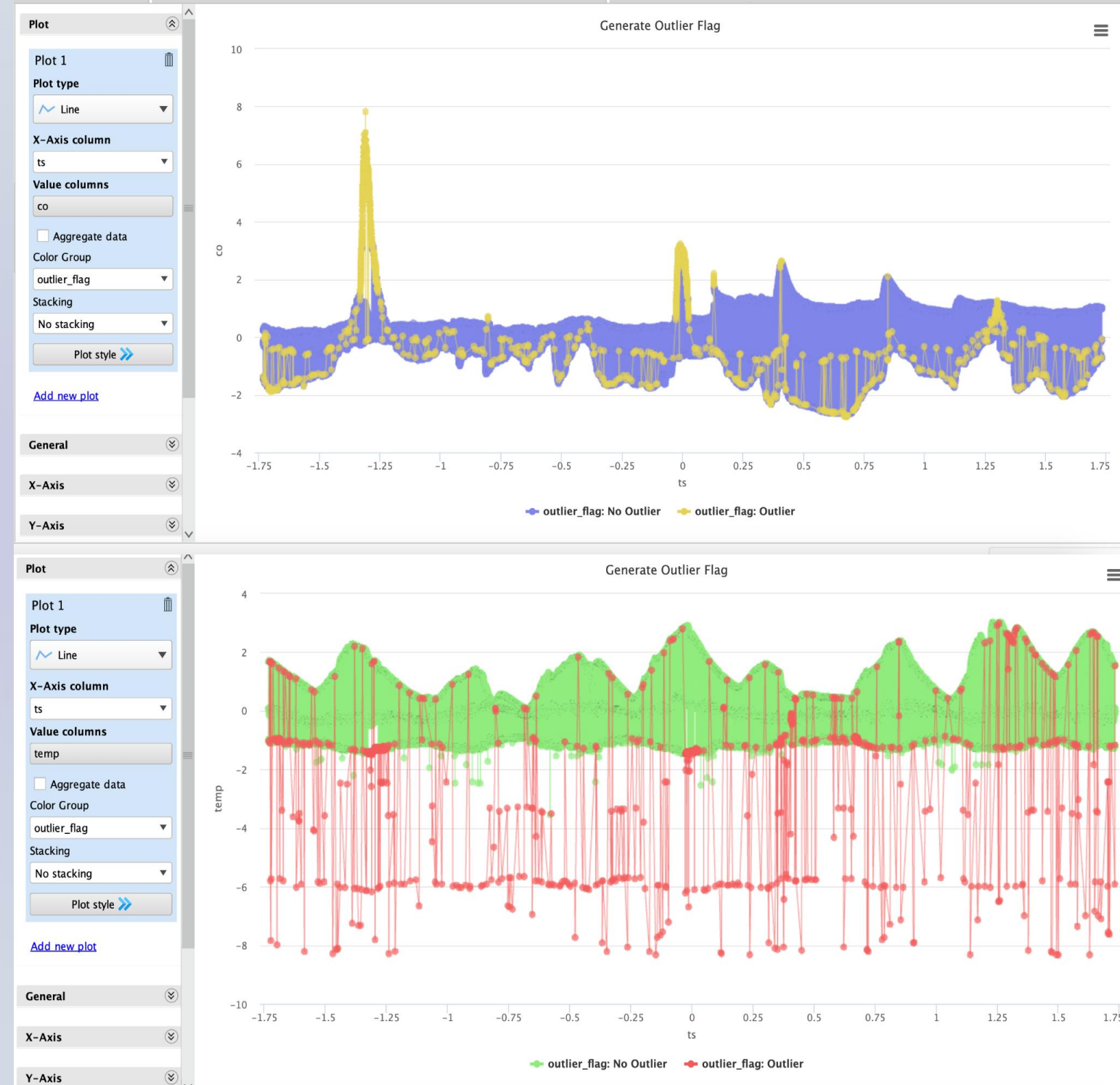


Диаграмма рассеяния для анализа аномалий

Выявление нетипичных комбинаций параметров

- Температура и влажность
- Газ и температура
- Другие комбинации параметров

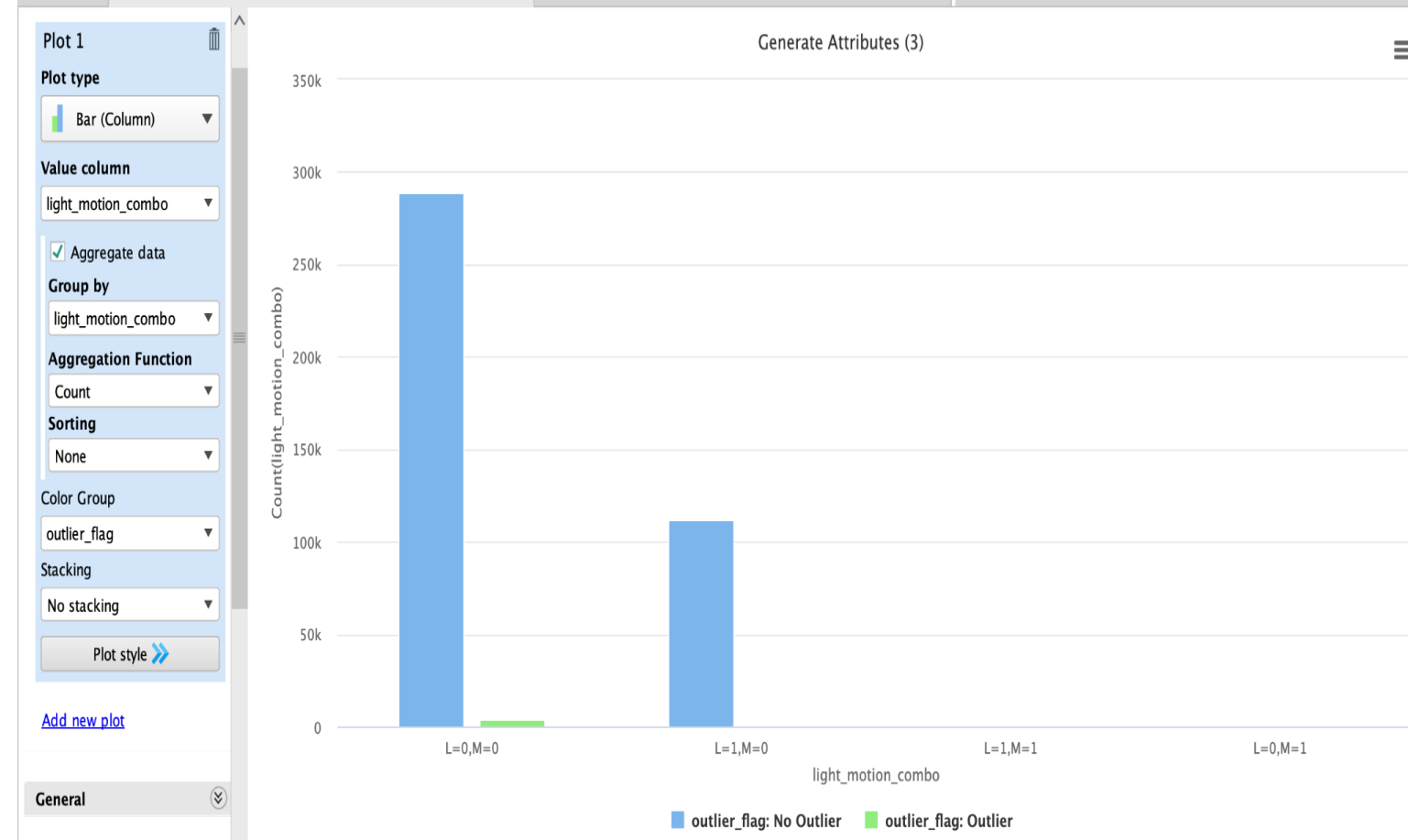
Диаграммы рассеяния позволяют
визуально определить точки,
выпадающие из общего
распределения данных





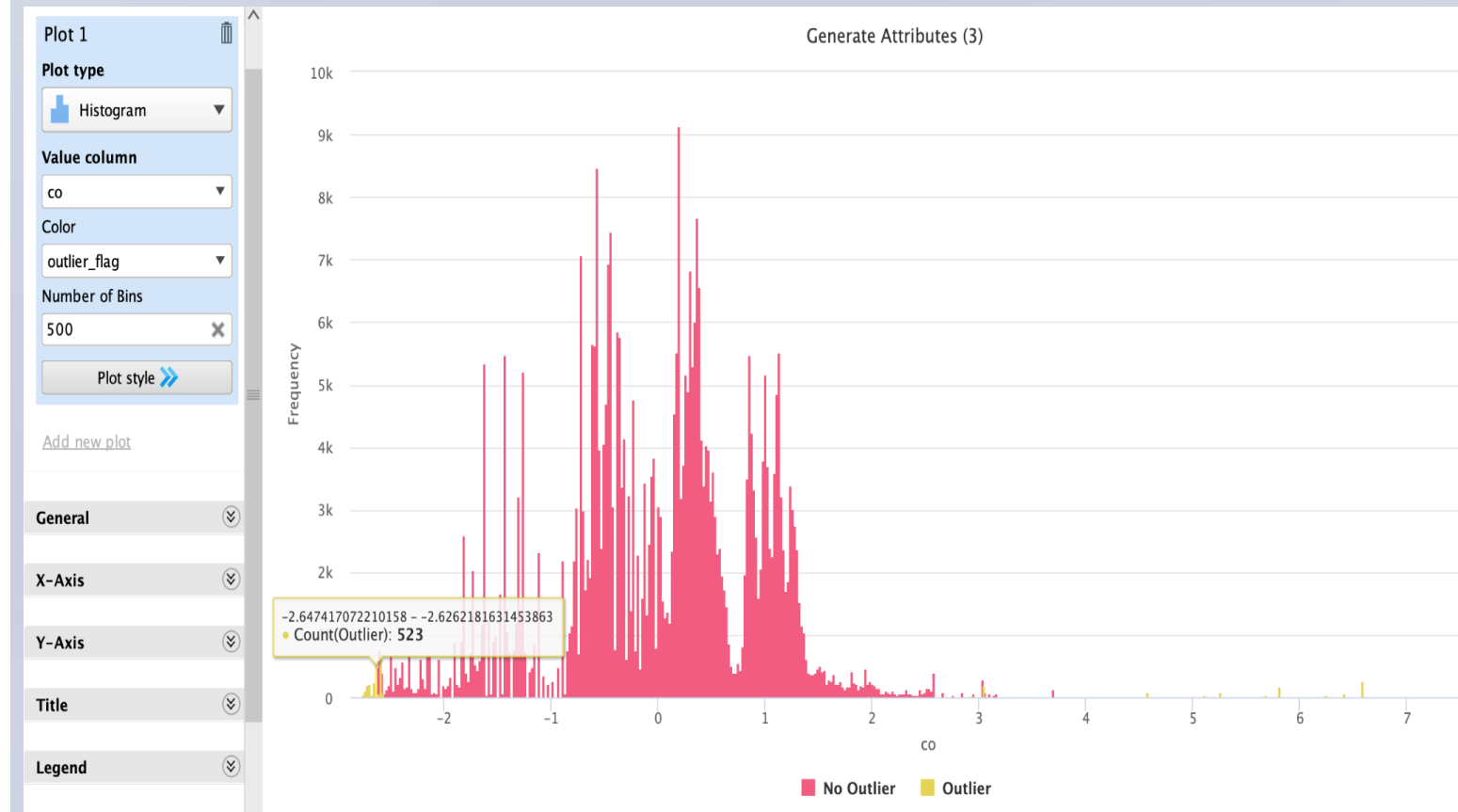
Анализ булевых признаков в IoT-данных

Аномальные комбинации булевых признаков (например, движение без света) выявляются через специальные графики и группировки



Гистограммы распределения аномалий

Гистограммы показывают, что большинство аномалий сосредоточено в экстремальных значениях признаков (например, концентрации CO).



Причины выявленных аномалий в IoT-данных



Неисправности датчиков

Физические поломки или сбои в работе измерительных устройств



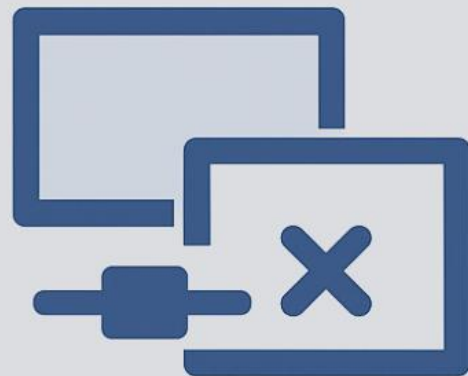
Сбои передачи данных

Проблемы с сетевым соединением или протоколами передачи



Реальные физические события

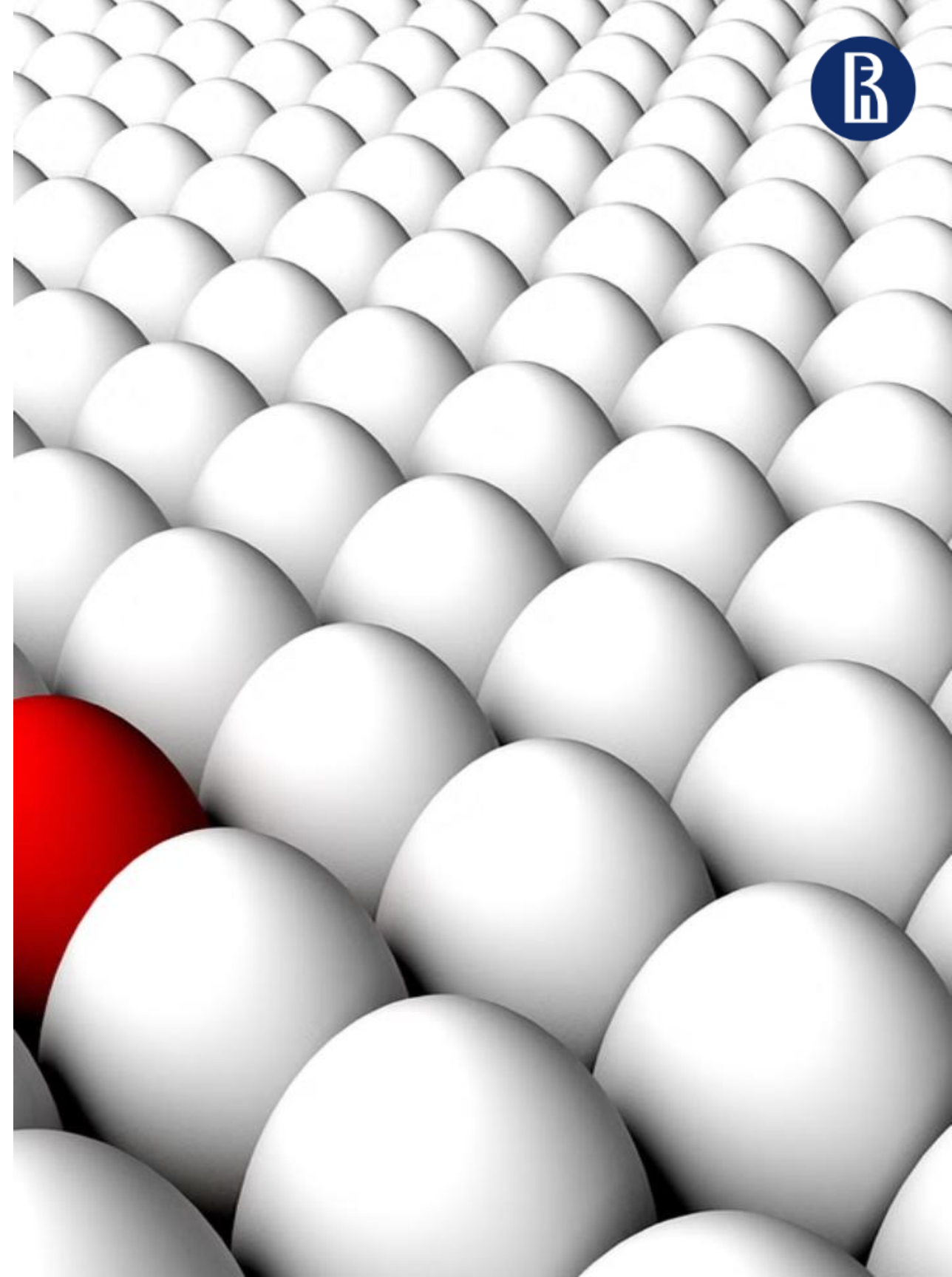
Утечки газа, экстремальные температуры и другие реальные аномалии



Ограничения анализа аномалий



Методы обнаружения аномалий могут приводить к ложным срабатываниям или пропускать скрытые аномалии. Эффективность сильно зависит от выбора параметров алгоритмов.



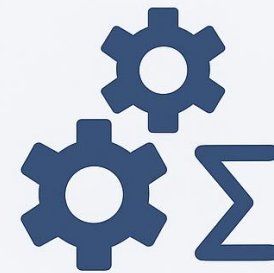
Рекомендации по анализу IoT-данных



Сочетание нескольких алгоритмов

Регулярная перекалибровка датчиков

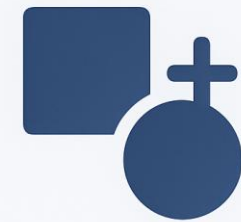
Проверка наиболее подозрительных аномалий вручную



Перспективные направления анализа аномалий



- Алгоритмы глубокого обучения
- Гибридные модели
- Подходы на основе ансамблей



Всё это обеспечивает более глубокий анализ многомерных данных.

Заключение по анализу аномалий



Проактивное управление
Предотвращение аварий до их возникновения благодаря раннему обнаружению отклонений.

Повышение надежности
Улучшение стабильности инфраструктуры через мониторинг скрытых проблем.

Оптимизация ресурсов
Сокращение затрат на обслуживание за счет точечного реагирования.

Улучшение процессов
Накопление данных об аномалиях помогает совершенствовать бизнес-процессы.

Грамотный анализ аномалий – ключ к устойчивости IoT-систем.