

**GloSIC'2020**

**2020 IEEE Global Smart Industry Conference**

Chelyabinsk, Russia, November 17–19, 2020

---

# **Cleaning Sensor Data in Smart Heating Control System**

Mikhail Zymbler, Yana Kraeva, Elizaveta Latypova,  
Sachin Kumar, Dmitry Shnayder, Alexander Basalaev

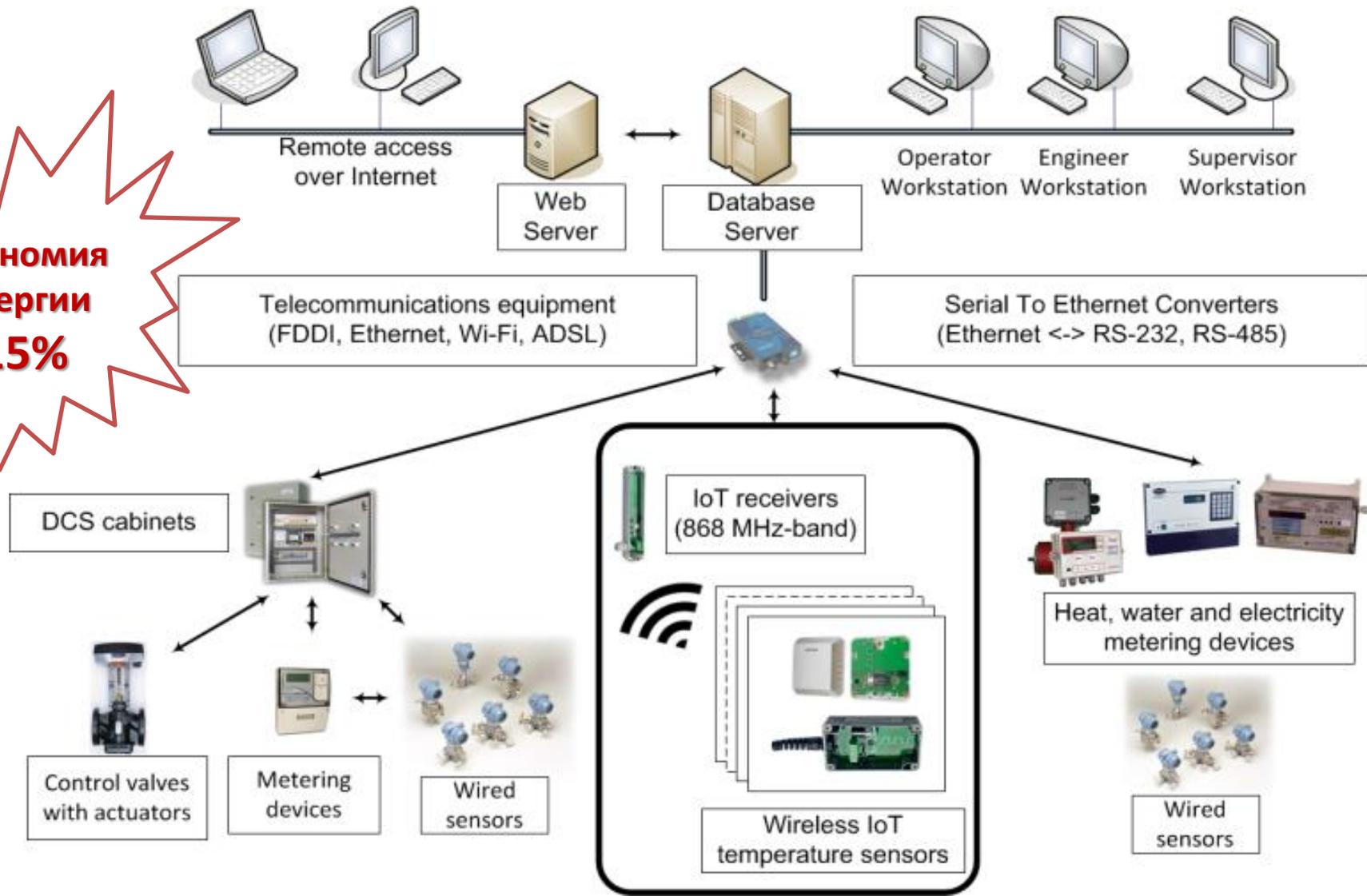
South Ural State University, Chelyabinsk, Russia

---

The study was financially supported by the Russian Foundation for Basic Research (grant No. 20-07-00140) and by the Ministry of Science and Higher Education of the Russian Federation (government order FENU-2020-0022 and contract no. 075-15-2019-1339 (14.578.21.0265, project identifier: RFMEFI57818X0265))

# Система умного управления отоплением в ЮУрГУ

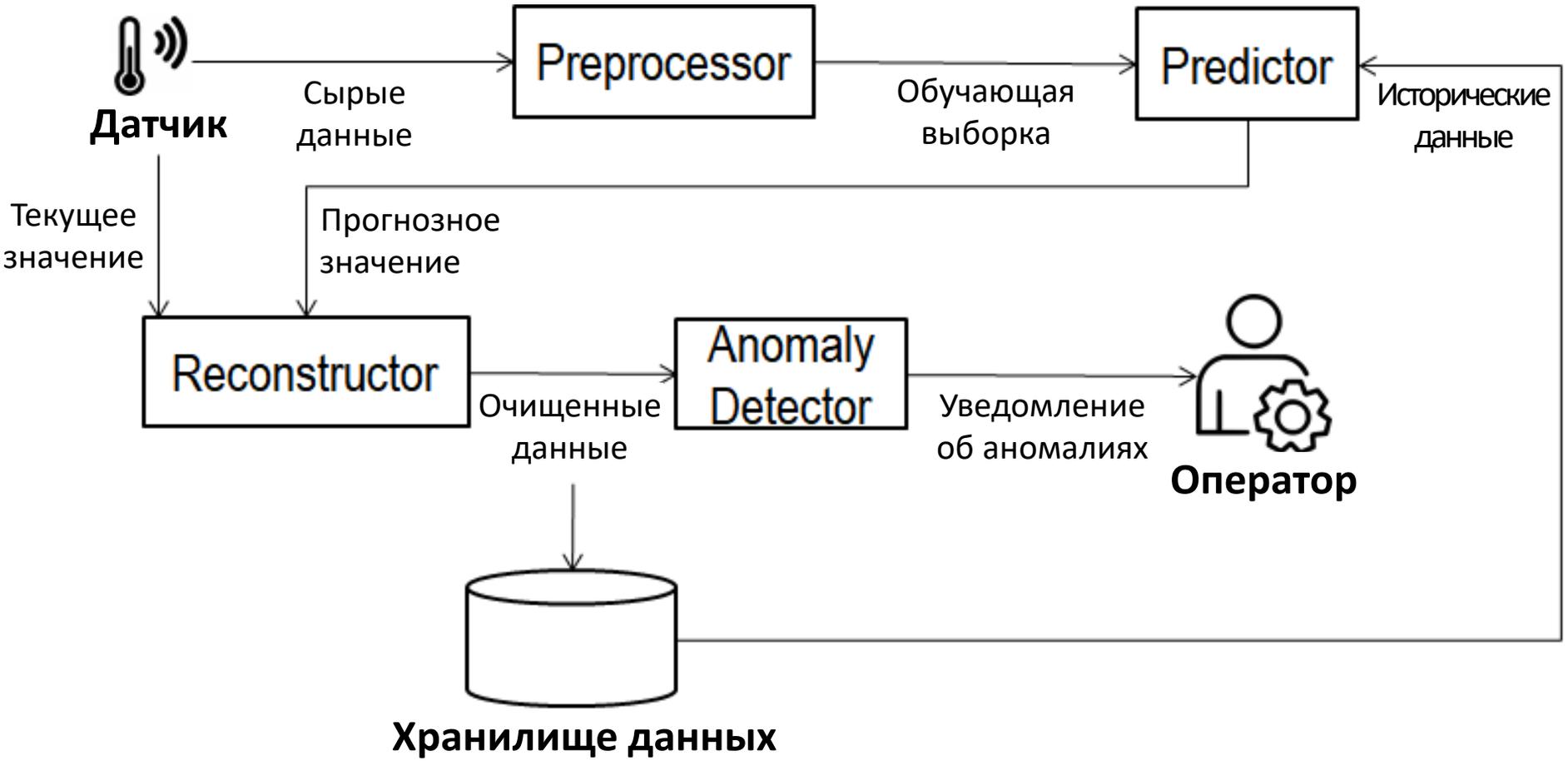
**Экономия  
энергии  
15%**



# Необходимость и основные функции очистки сенсорных данных

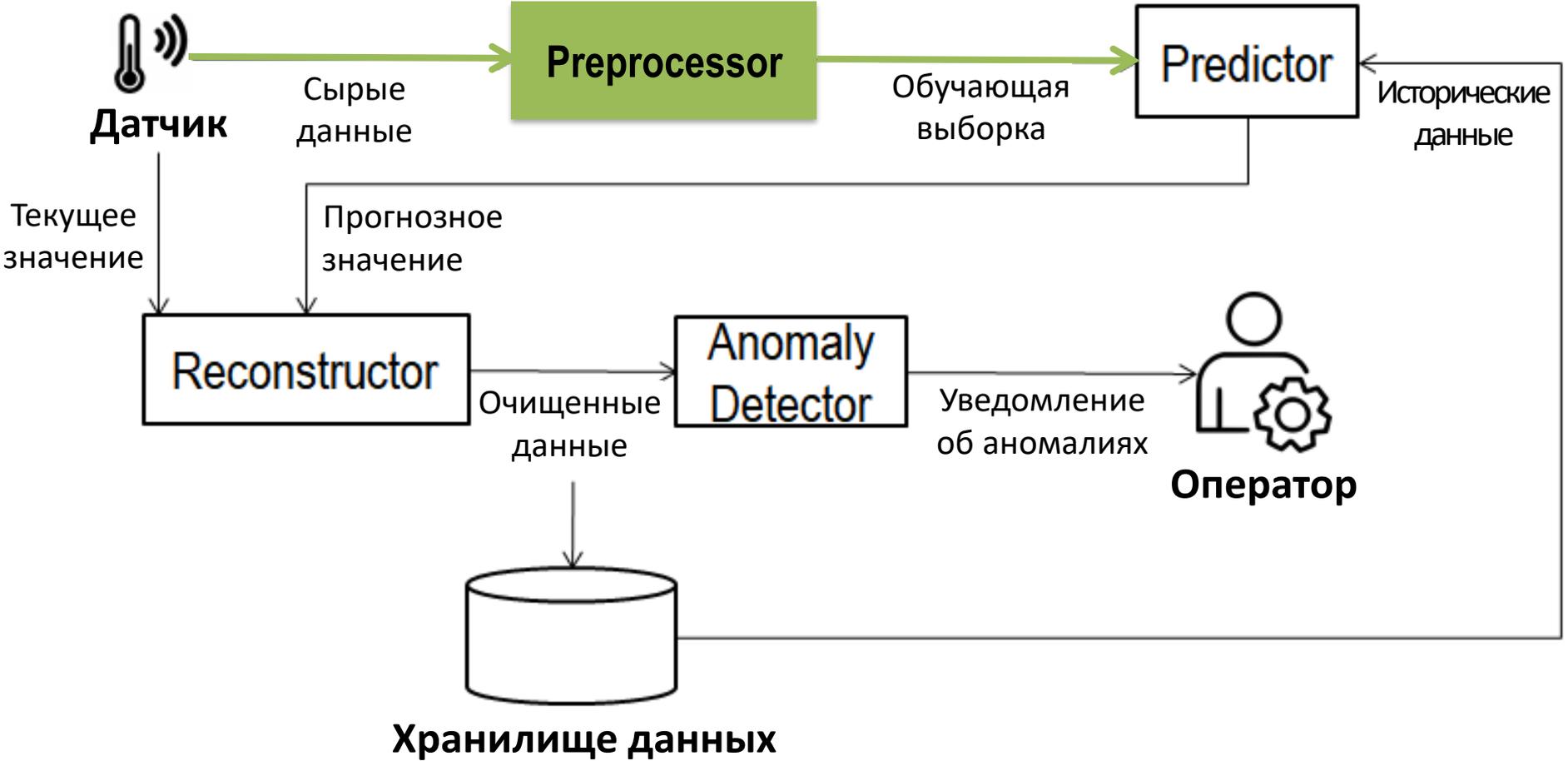
- Последствия выбросов и пропусков в данных
  - сбои в работе датчика и инженерной сети
  - повышенные затраты энергии
  - некорректное управление отоплением в автоматическом режиме
- Основные функции модуля очистки
  - оперативная замена пропусков и выбросов на правдоподобные значения
  - оперативное обнаружение аномалий для уведомления оператора

# Модуль очистки данных: архитектура

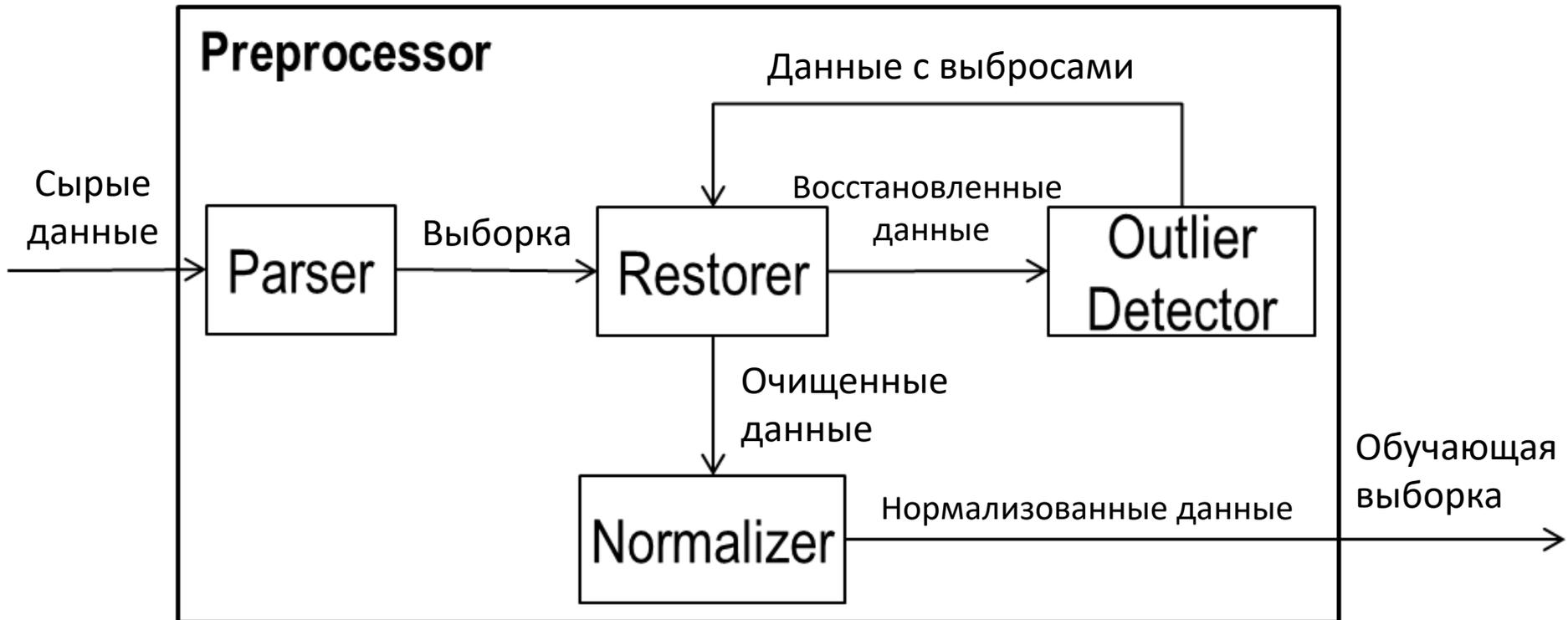


# Модуль очистки данных: Preprocessor

## Подготовка обучающей выборки для нейронной сети подсистемы Predictor

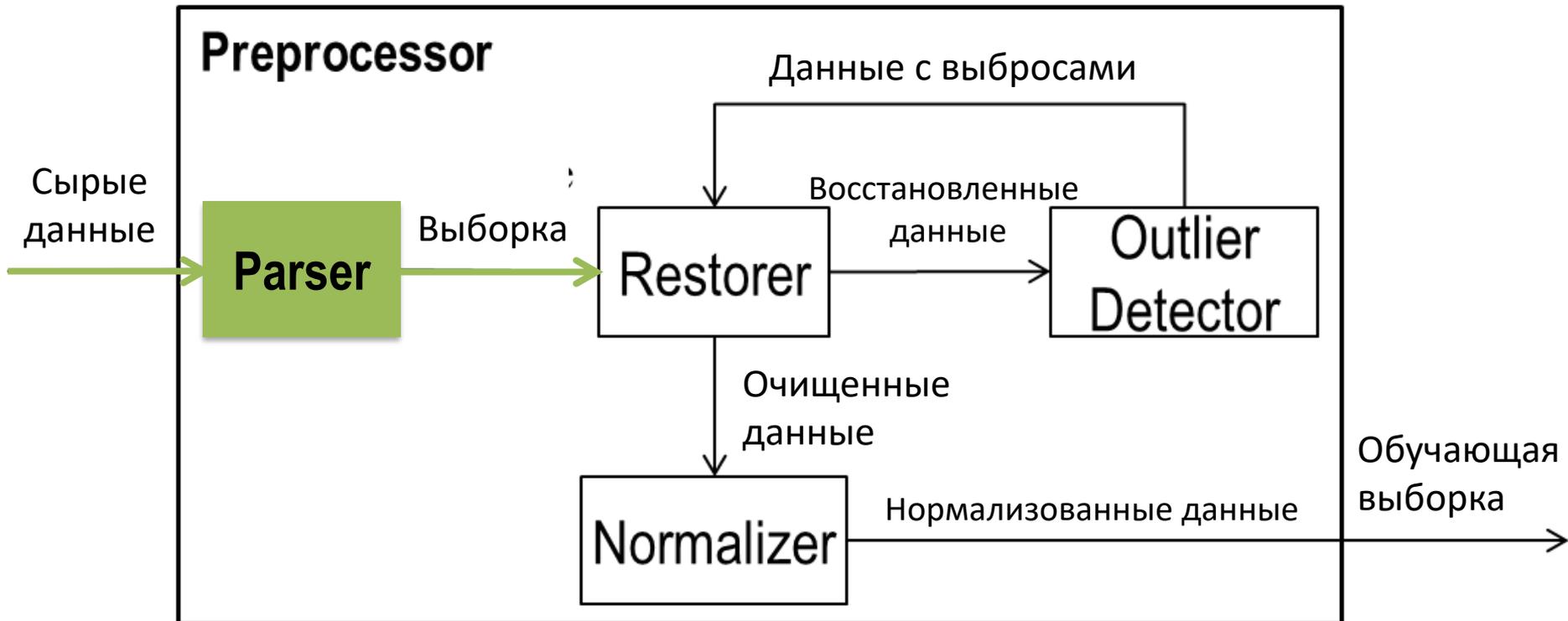


# Preprocessor: структура

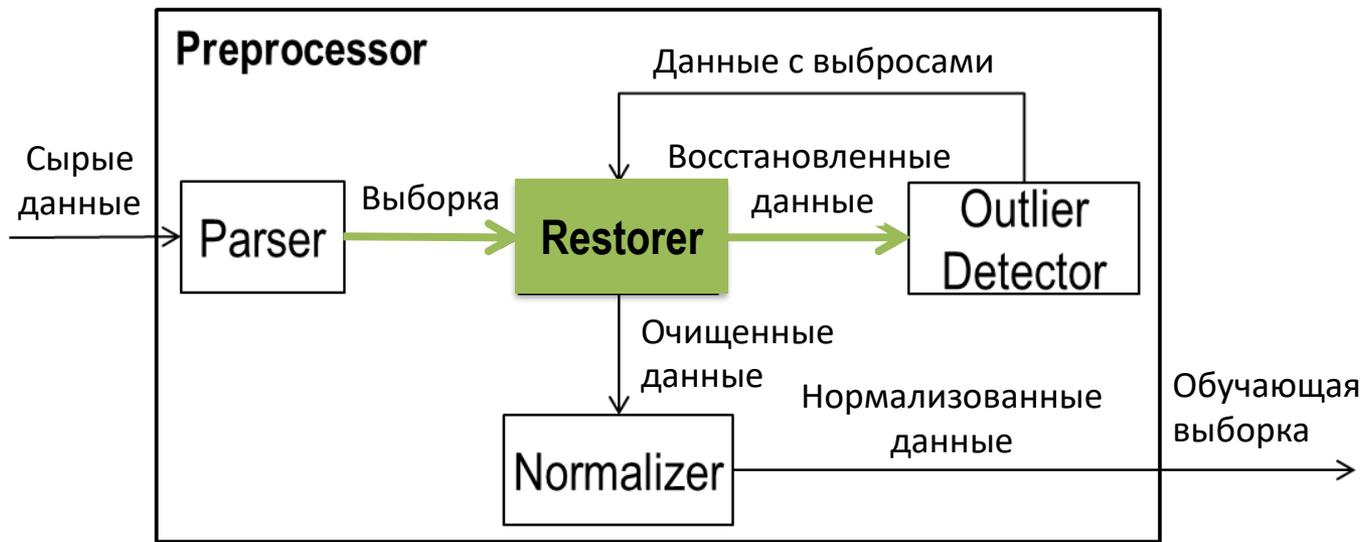


# Preprocessor: модуль Parser

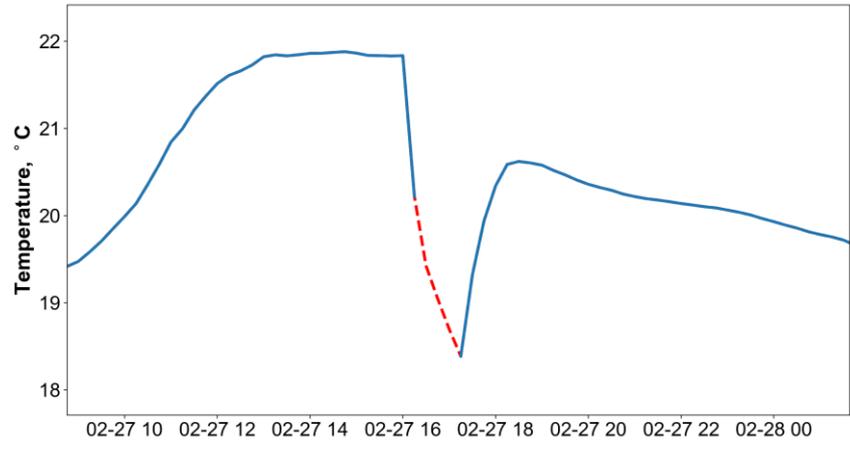
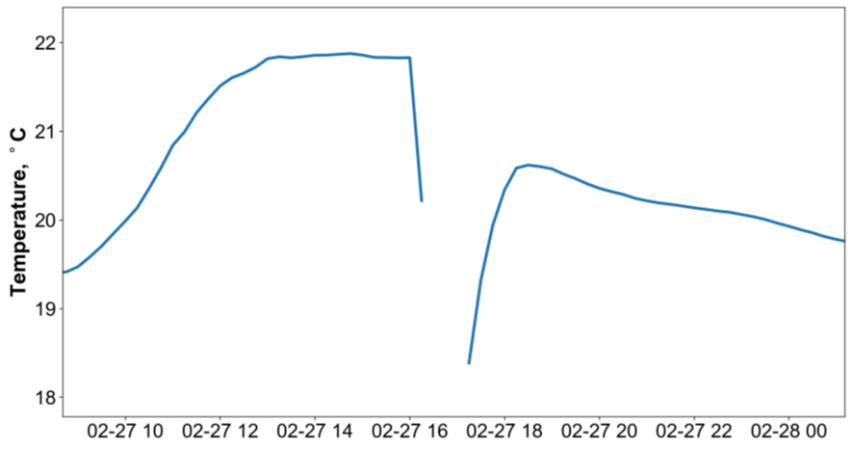
Извлечение сырых показаний датчика из хранилища данных и их предобработка



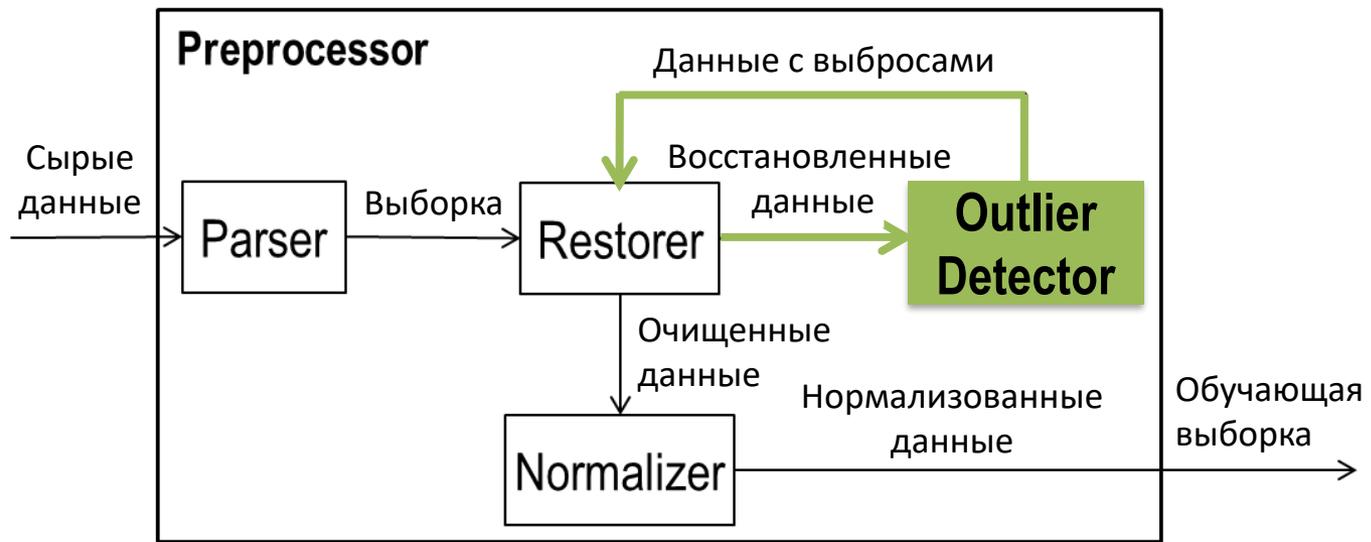
# Preprocessor: модуль Restorer



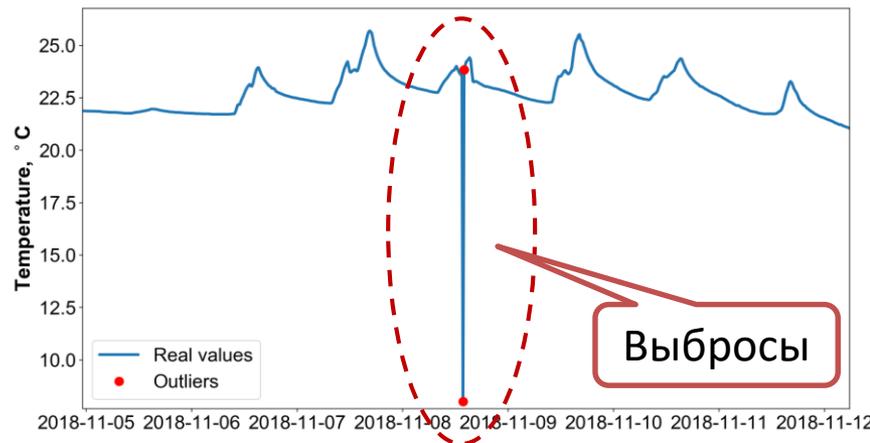
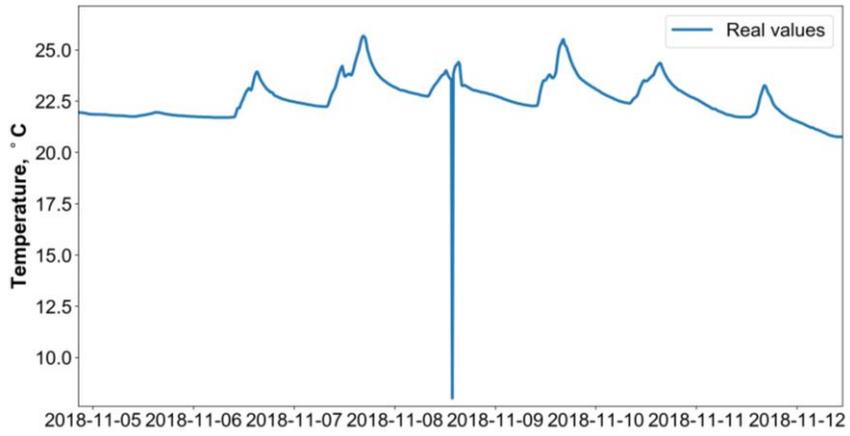
## Замена пропусков на правдоподобные значения



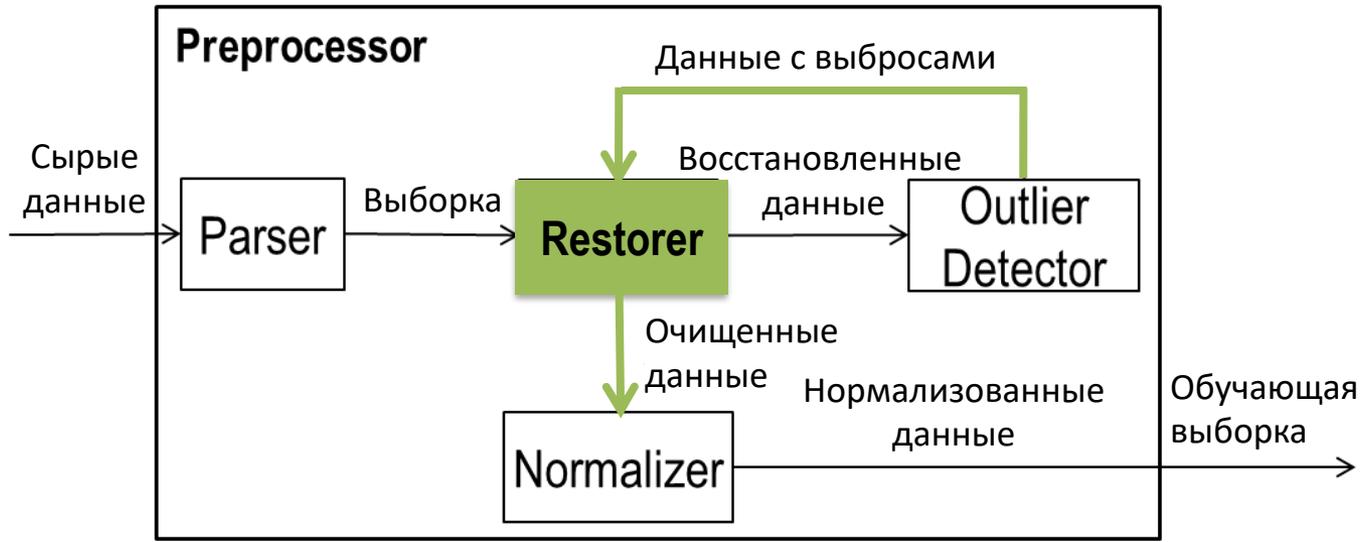
# Preprocessor: модуль Outlier Detector



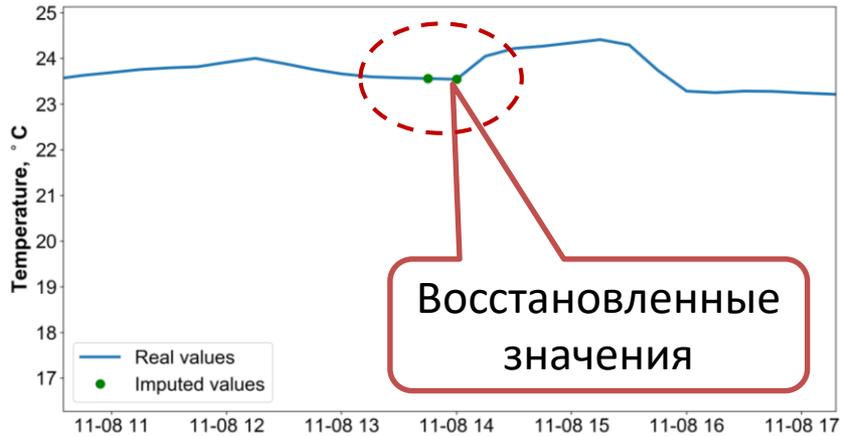
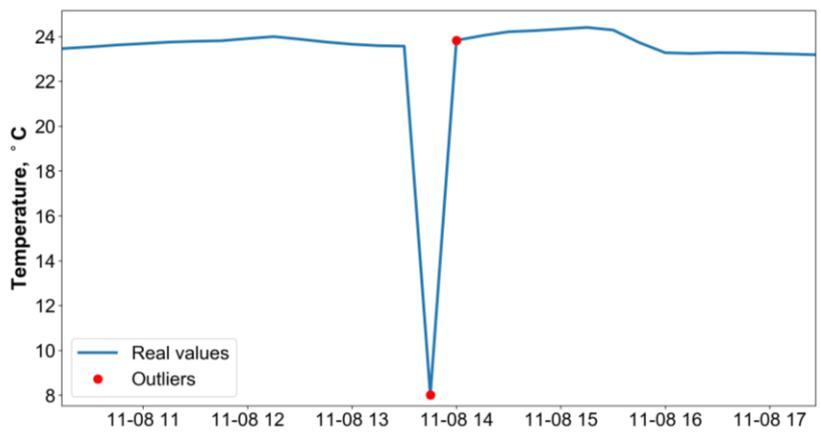
## Нахождение точечных выбросов



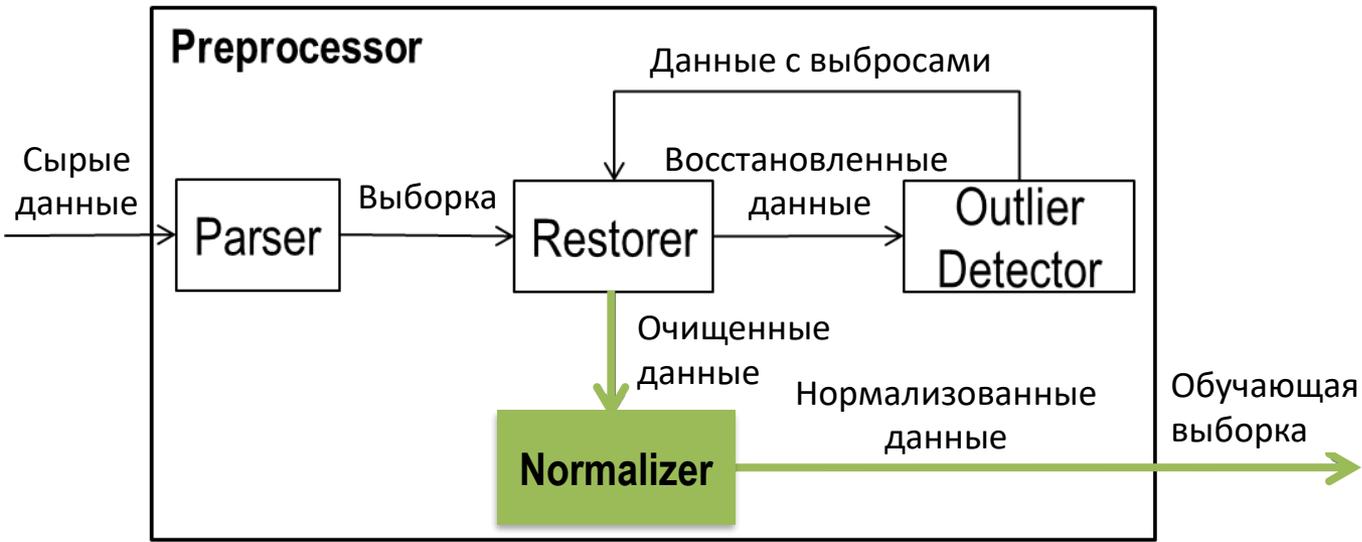
# Preprocessor: модуль Restorer



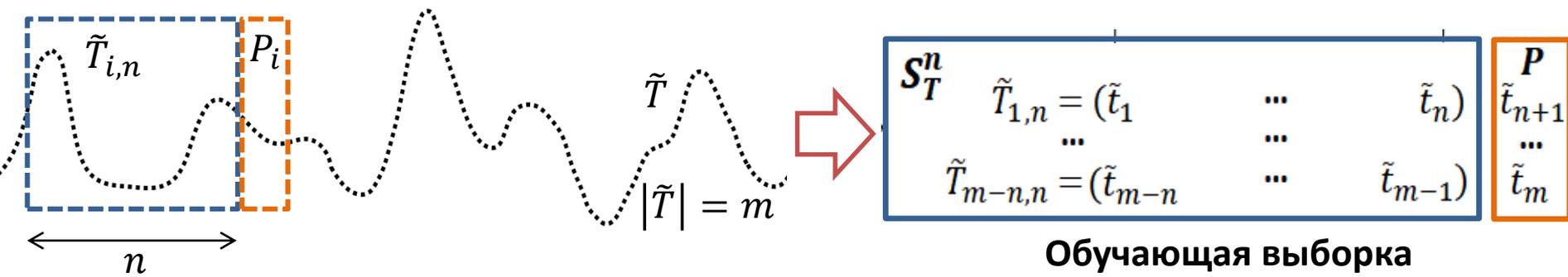
## Замена выбросов на правдоподобные значения



# Preprocessor: модуль Normalizer

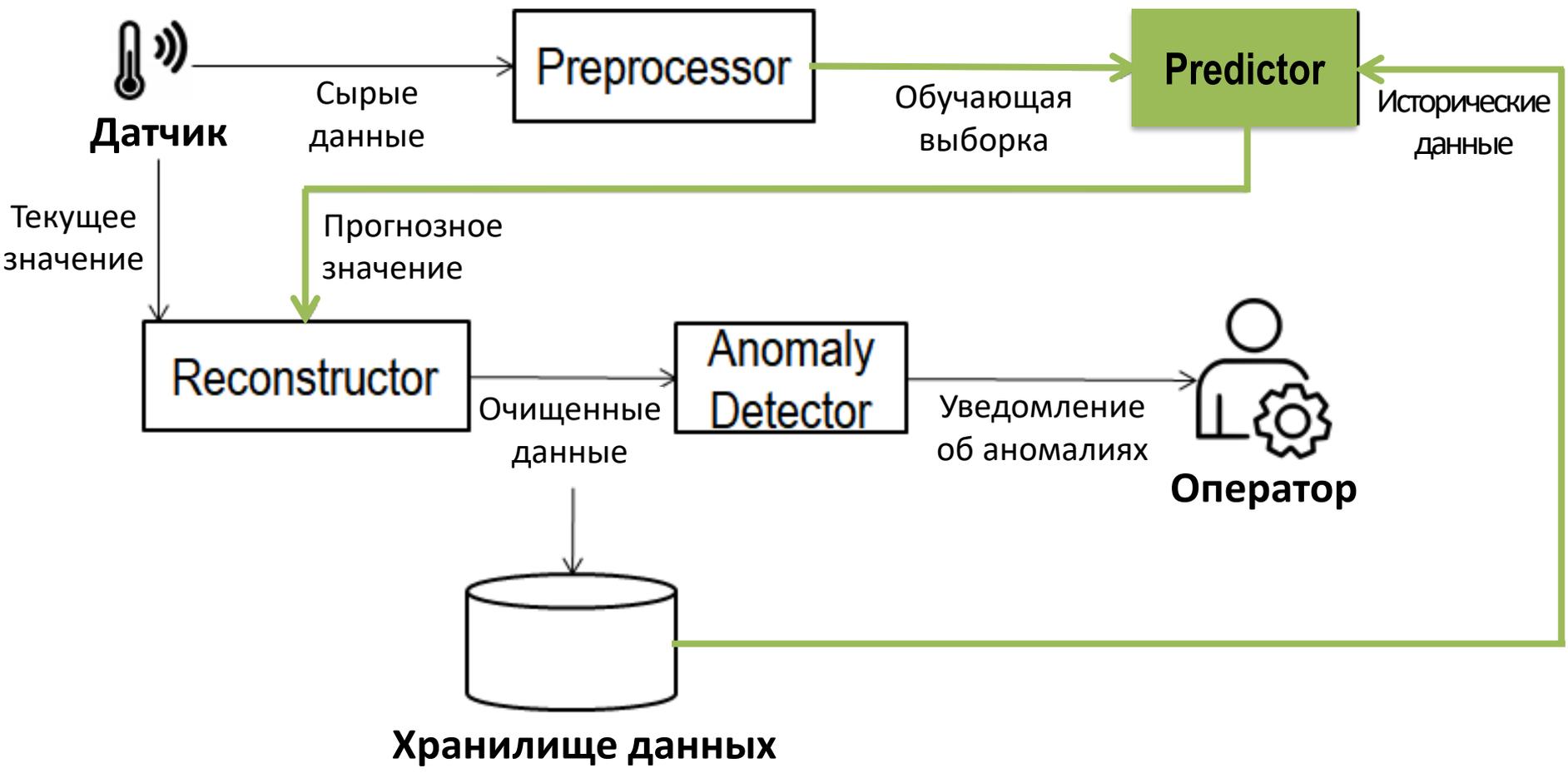


Min-max нормализация данных  $\tilde{t}_i = \frac{t_i - t_{min}}{t_{max} - t_{min}}, \tilde{t}_i \in [0; 1]$

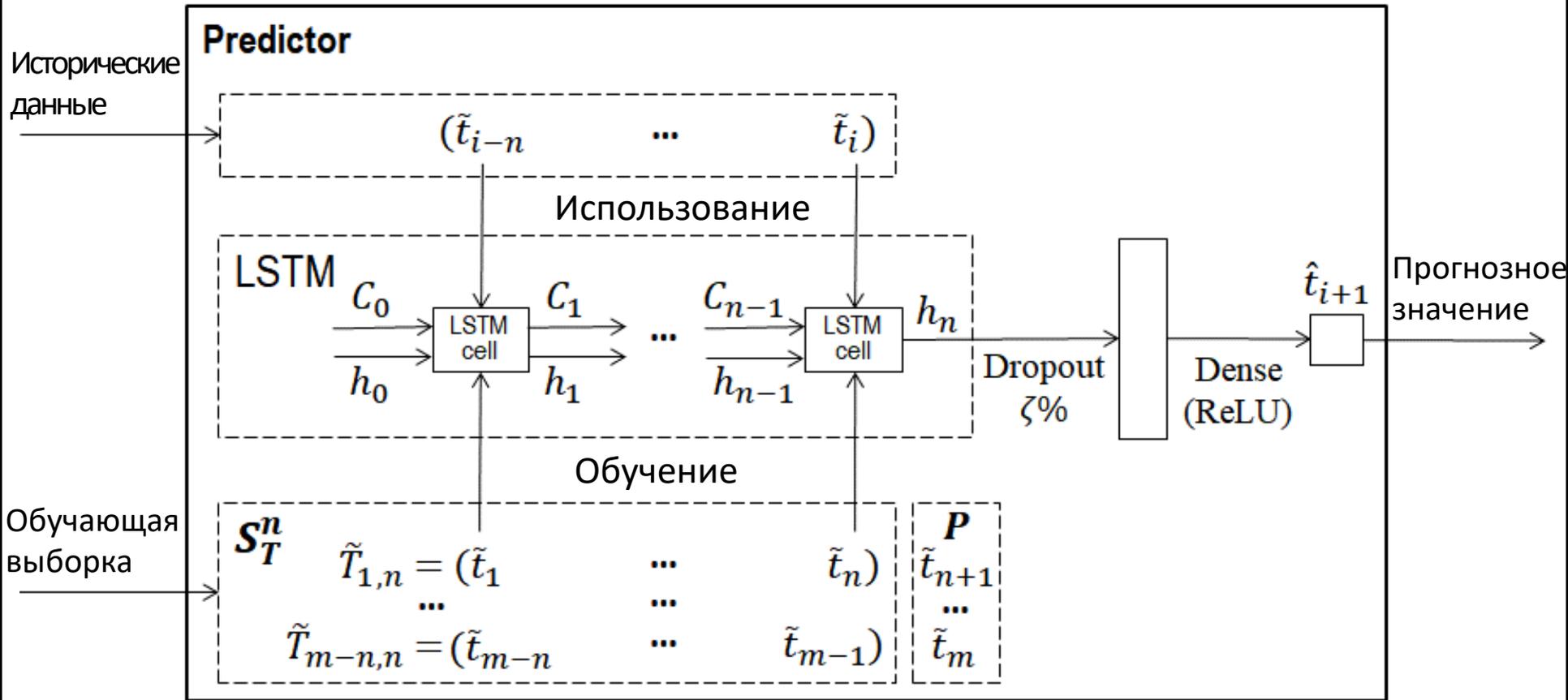


# Модуль очистки данных: Predictor

## Нейронная сеть для прогноза показаний датчика

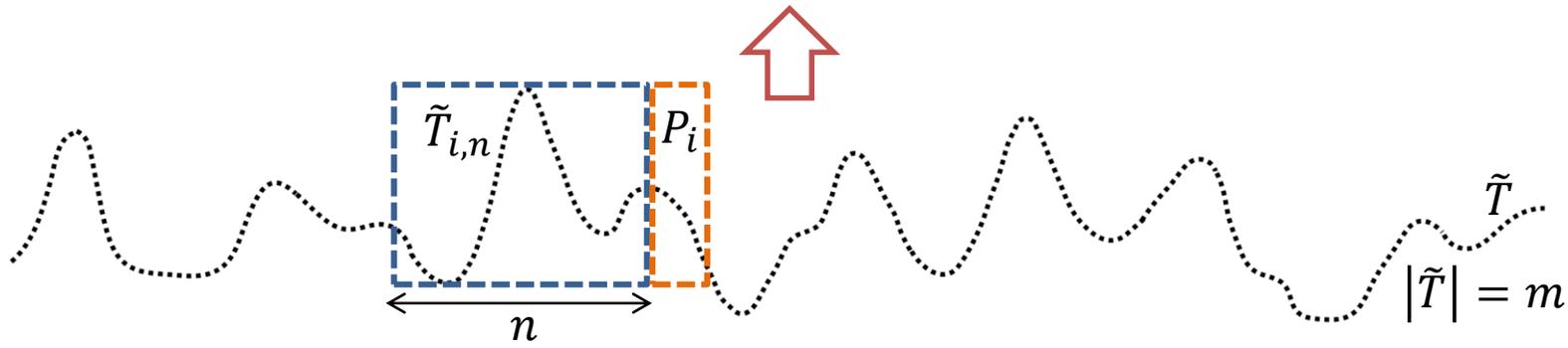
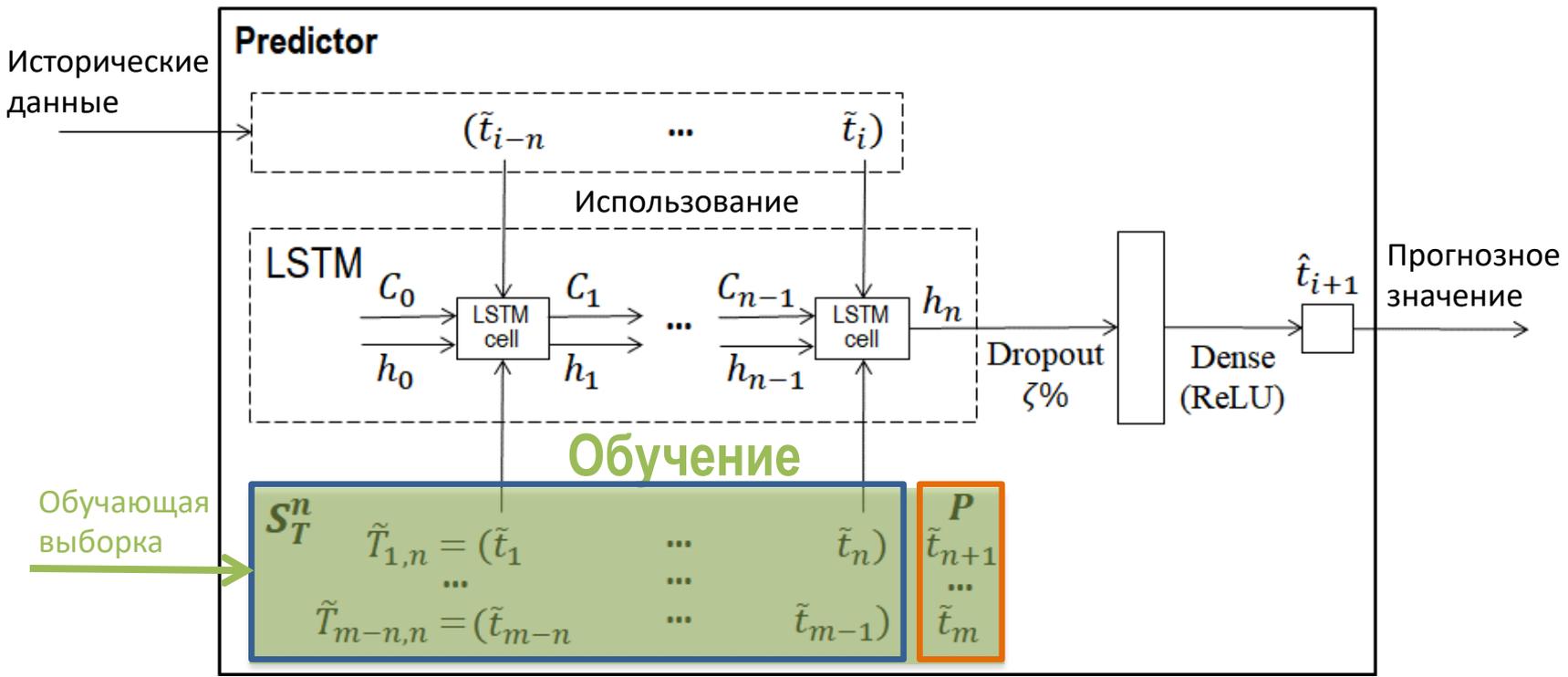


# Predictor: структура

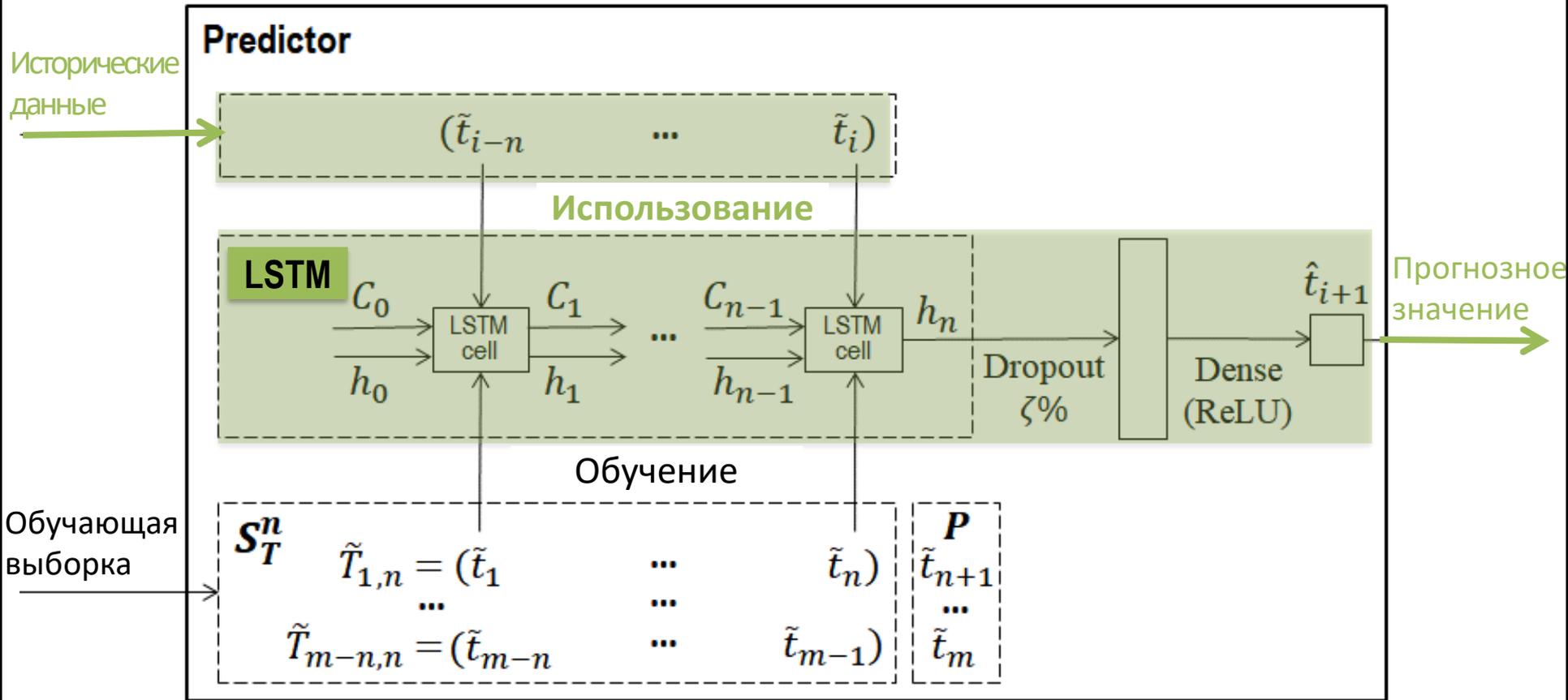




# Predictor: обучение нейросети

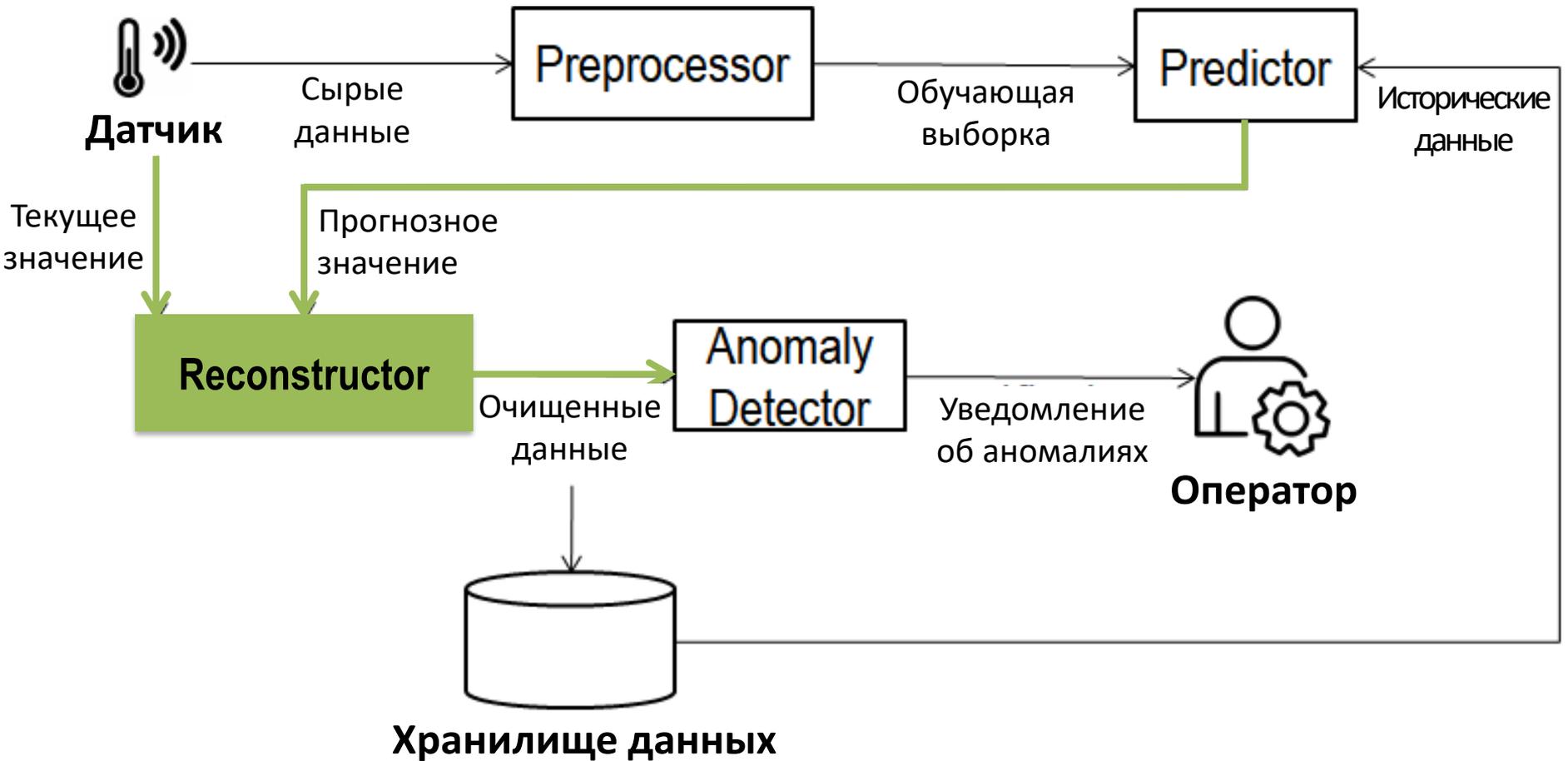


# Predictor: использование



# Модуль очистки данных: Reconstructor

## Замена выбросов и пропусков на правдоподобные значения



# Reconstructor

- Идея

**ЕСЛИ**  $|t_{current} - t_{predicted}| \geq \varepsilon$  **ТО**

$t_{current}$  считаем выбросом

$t_{current} := t_{predicted}$

- Реализация: **правило трех сигм**

Порог несхожести  $\varepsilon := \mu + k\sigma$ ,

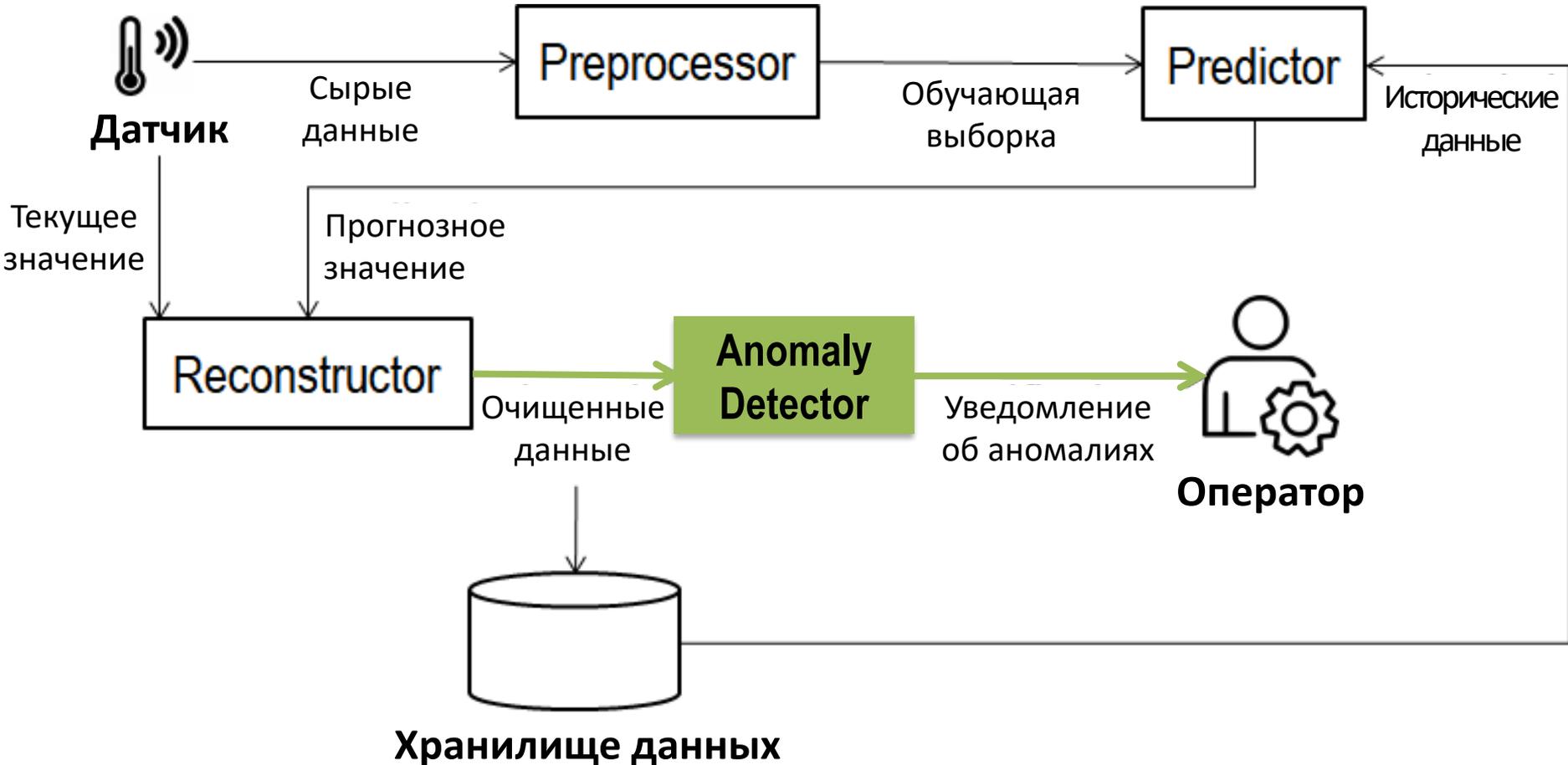
$\mu$  – среднее значение ошибок прогнозирования,

$\sigma$  – стандартное отклонение ошибок прогнозирования,

$k$  – параметр (типично  $k = 3$ )

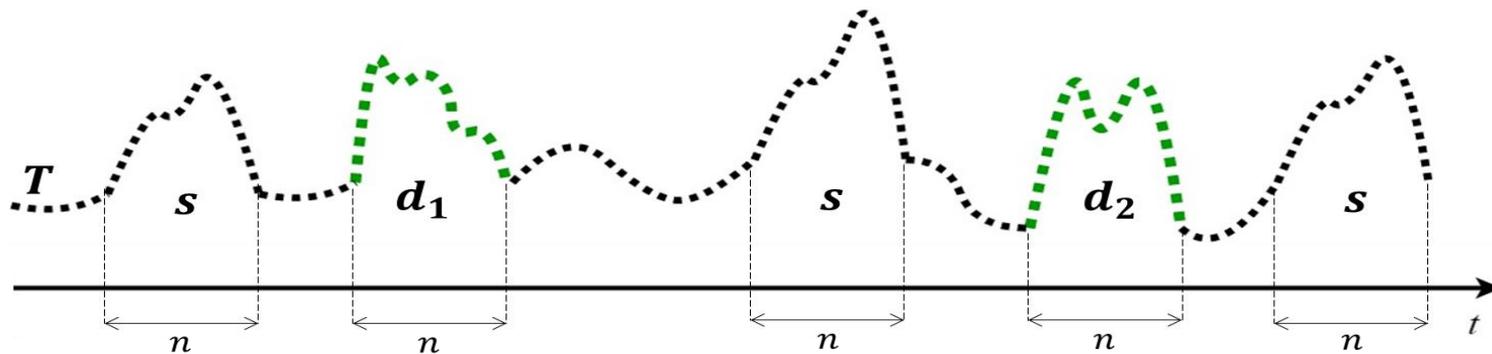
# Модуль очистки данных: Anomaly Detector

## Выявление аномалий в показаниях датчика



# Anomaly Detector

- Идея
  - поиск аномальных подпоследовательностей, оканчивающихся текущим показанием температурного датчика
  - длина аномалии соответствует значимым промежуткам: 0.5 суток, 1 суток, 2 суток
- Реализация: **диссонанс временного ряда**
  - Диссонанс – это подпоследовательность с Евклидовым расстоянием не менее  $r$  до ближайшего соседа ( $r$  – параметр)
  - $\mathcal{D} = \{d_1, d_2, \dots\}$ ,  $d_i \in \mathcal{D} \Leftrightarrow \min_{s \in T, s \cap d_i = \emptyset} \text{ED}(d_i, s) \geq r$



# Эксперименты

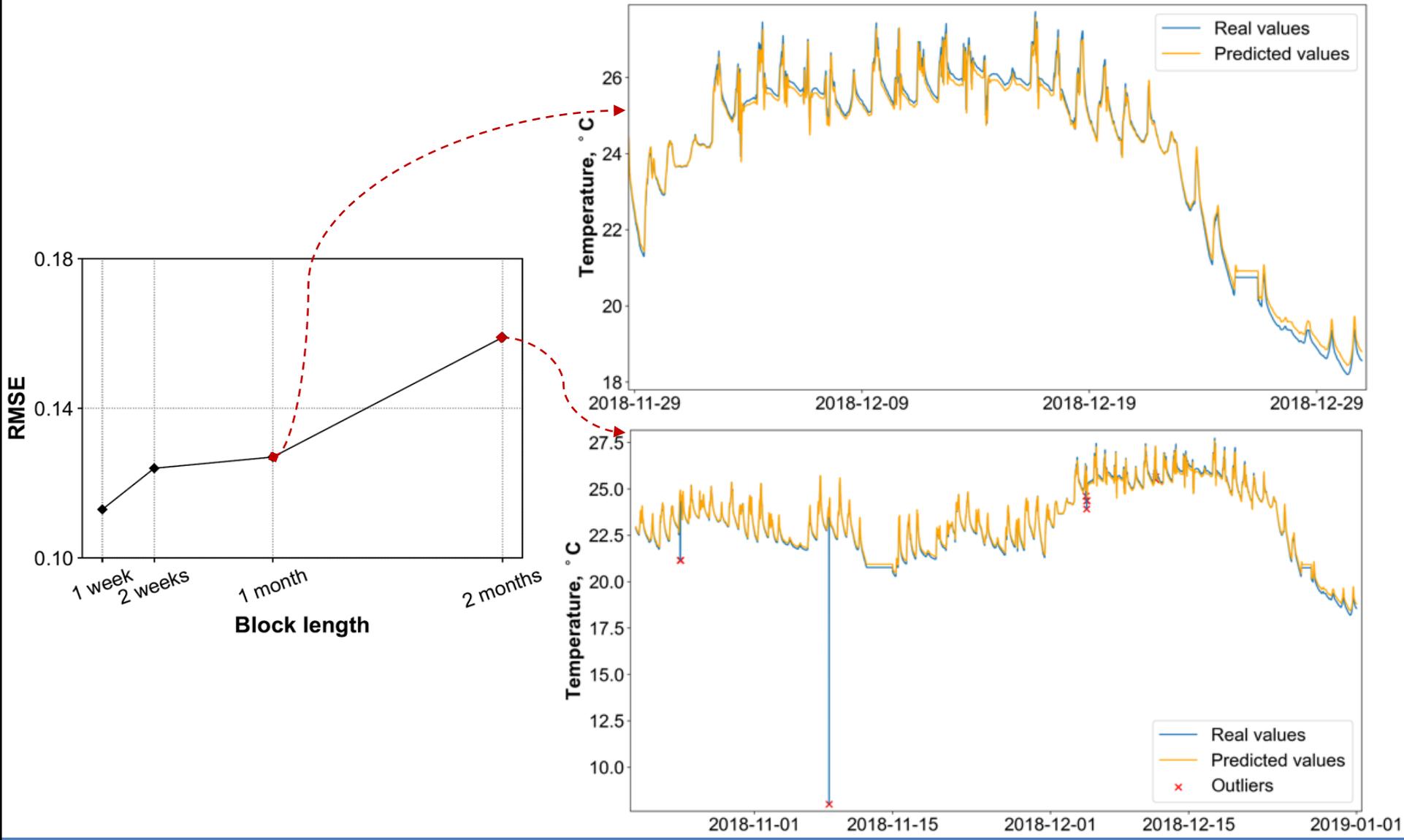
- Цель: оценка точности соответствия реальных и синтезированных показаний датчика
- Данные: январь-декабрь 2018 г., датчик в ауд. 434/36
- Мера:

$$\text{RMSE} = \sqrt{\frac{1}{|B|} \sum_{i=1}^{|B|} (t_i - \hat{t}_i)^2}$$

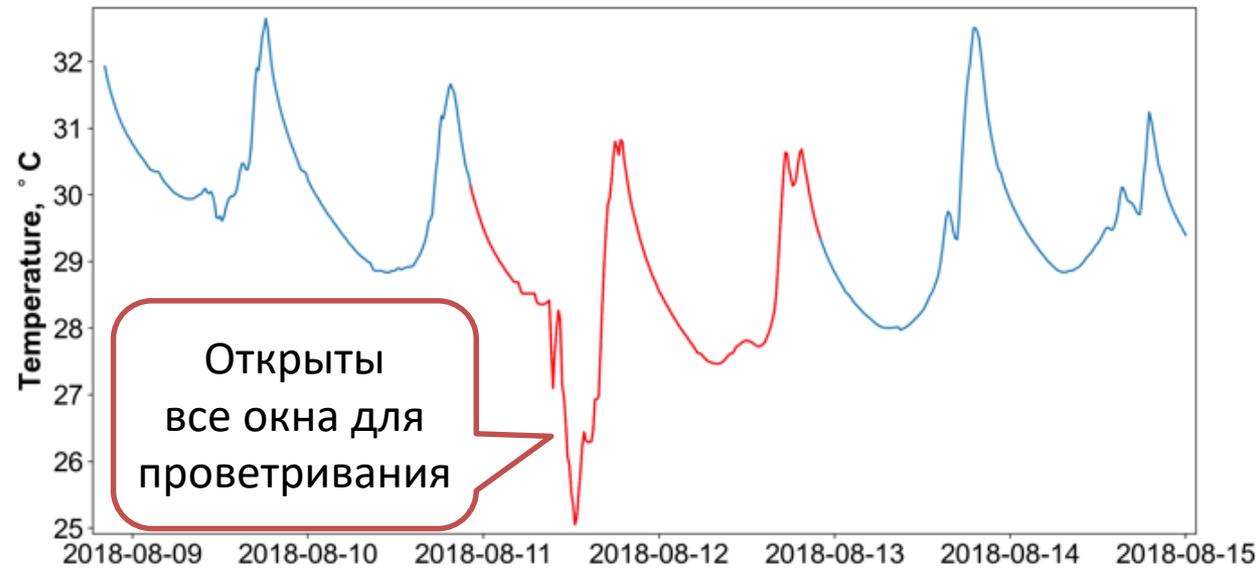
$t_i$  и  $\hat{t}_i$  – реальное и синтетическое значения датчика,  
 $|B|$  – горизонт восстановления

- Параметры модели:
  - длина окна: 48 точек (12 часов работы датчика)
  - длина скрытого состояния RNN: 32
  - функция потерь: MSE (mean square error)
  - оптимизатор: Adam
  - # эпох и размер подвыборки: 15 и 32

# Результаты экспериментов: точность



# Результаты экспериментов: найденные аномалии



# Спасибо за внимание!

## Вопросы?

**Яна Александровна Краева**

Отдел интеллектуального анализа данных  
и виртуализации Лаборатории  
суперкомпьютерного моделирования,  
программист