



Факультет вычислительной математики и кибернетики
МГУ имени М.В. Ломоносова

Программа профессиональной переподготовки
«Машинное обучение и управление большими
данными»

Разработка сервиса определения работоспособности приборов учета
теплоснабжения с применением методов машинного обучения

Слушатель: Макарычев Николай Владимирович

Научный руководитель: Мочалова Юлия Дмитриевна

Москва, 2026 г.

Актуальность разработки сервиса диагностики ПУ

ПРОБЛЕМА:

Достоверный учёт тепловой энергии — основа финансовых взаиморасчётов в сфере ЖКХ. Определить, исправен ли прибор учёта или нет, необходимо как можно раньше — это и есть задача **мониторинга**. Сегодня такой мониторинг осуществляется **вручную**: инженеры и аналитики самостоятельно просматривают каждую ведомость потребления.

Ручной мониторинг



3 критических ограничения

1. Высокие трудозатраты

Обработка растущего объёма данных требует пропорционального роста численности аналитиков

2. Субъективность решений

Качество диагностики зависит от квалификации и внимательности конкретного специалиста

3. Пропуск скрытых дефектов

Человек не способен выявлять сложные многомерные аномалии в показаниях

Пропуск неисправности → **Переход на нормативы** → **Счета за тепло в разы выше фактического потребления**

Вывод: Требуется автоматизированный сервис диагностики на основе **машинного обучения**, который обеспечит:

Исключение

человеческого фактора при принятии решений — решения принимаются моделью объективно и единообразно, без зависимости от квалификации специалиста

Снижение

трудозатрат на анализ ведомостей — автоматизация сокращает время обработки данных и освобождает специалистов от рутинной работы

Максимальное

обнаружение неисправных приборов учёта — за счёт оптимизированного порога модель обеспечивает высокую полноту обнаружения, минимизируя финансовые риски

Доступность

не требует глубоких экспертных знаний — достаточно загрузить ведомости и опираться на готовое заключение сервиса

Цели и задачи

ЦЕЛЬ:

Разработать локальный сервис для автоматической диагностики приборов учета тепла на основе методов машинного обучения, обеспечивающий **максимальную полноту обнаружения** неисправностей при приемлемом уровне ложных срабатываний

ЗАДАЧИ:

1

Анализ предметной области

Изучить типы приборов учета, их диагностические параметры и классифицировать возможные неисправности и аномалии



2

Разработка ML-модели

Провести разведочный анализ данных, обучить и сравнить модели классификации, выбрать лучшую и оптимизировать порог принятия решений



3

Разработка сервиса

Спроектировать клиент-серверную архитектуру, реализовать бэкенд (FastAPI), фронтенд (Streamlit) и модуль экспорта отчётов



4

Тестирование и экономика

Провести тестирование на исторических данных и оценить экономическую эффективность внедрения сервиса в деятельность теплоснабжающих организаций

Посуточная ведомость показаний прибора учета

отопления (вентиляции) за март 2026 г. (с 01.03.2026 по 31.03.2026)

Потребитель: ГБУ "Норильские горы М"

ЦТП №: 35-01-4108/255

Адрес: ул. Московского Государственного Университета, д. 1

Телефон: Не указан

Прибор: —

Сер.номер: 25011755

Расход под.: 0,1-100 м³/ч Ду" мм

Модель: АТ-Т-2

Версия ПО: 1.05

Расход обр.: 0,100-100 м³/ч Ду" мм

Дата	Q [Гкал]	Gпод [т]	Gобр [т]	Gпод-Gобр		tпод [°C]	tобр [°C]	Rпод [ат]	Rобр [ат]	Tнар [°C]	Классиф. ошибок
				Подмес	Утечка						
01.03.26	3.63	500.36	498.18		2.176	54.3	47.1	6.58	4.63	24.00	
02.03.26	3.77	500.30	498.17		2.131	55.6	48.0	6.66	4.71	24.00	
03.03.26	4.00	500.15	497.96		2.189	57.3	49.3	6.73	4.80	24.00	
04.03.26	4.22	499.86	497.67		2.191	59.0	50.6	6.78	4.84	24.00	
05.03.26	3.91	503.79	501.67		2.111	56.7	48.9	6.79	4.83	24.00	
06.03.26	4.44	505.98	503.88		2.098	60.4	51.6	7.19	5.22	24.00	
07.03.26	4.20	506.18	503.98		2.199	58.6	50.4	6.99	5.00	24.00	
08.03.26	4.07	506.20	504.01		2.188	57.5	49.5	6.99	5.01	24.00	
09.03.26	4.58	505.82	503.64		2.180	61.6	52.5	7.18	5.19	24.00	
10.03.26	3.00	503.61	501.43		2.188	49.9	44.0	6.81	4.85	24.00	
11.03.26	2.82	501.66	499.44		2.215	48.6	43.0	7.02	5.07	24.00	
12.03.26	2.68	501.80	499.62		2.176	47.7	42.4	6.93	4.99	24.00	
13.03.26	2.61	502.17	499.97		2.207	47.3	42.1	6.84	4.90	24.00	
14.03.26	2.70	502.28	500.09		2.184	48.0	42.6	6.78	4.83	24.00	
15.03.26	2.93	504.91	502.75		2.160	49.7	44.0	6.83	4.87	24.00	
16.03.26	3.15	507.12	504.93		2.191	51.5	45.3	6.91	4.93	24.00	
17.03.26	3.05	507.14	504.93		2.215	50.9	44.9	7.02	5.06	24.00	
18.03.26	2.78	507.38	505.07		2.309	48.7	43.2	7.10	5.13	24.00	
19.03.26	2.77	507.39	505.11		2.277	48.6	43.2	7.07	5.10	24.00	
20.03.26	2.97	503.97	501.78		2.191	50.5	44.6	7.14	5.20	24.00	
21.03.26	3.00	501.62	499.45		2.176	50.6	44.6	7.05	5.13	24.00	
22.03.26	2.63	501.91	499.77		2.141	47.9	42.6	6.78	4.86	24.00	
23.03.26	2.54	501.74	499.57		2.168	47.1	42.1	6.66	4.74	24.00	
24.03.26	2.67	501.63	499.43		2.203	48.2	42.9	6.92	5.00	24.00	
25.03.26	2.47	505.20	502.99		2.207	46.7	41.8	7.06	5.11	24.00	
26.03.26	2.73	507.40	505.18		2.215	48.7	43.3	7.16	5.19	24.00	
27.03.26	2.38	507.56	505.36		2.199	46.2	41.5	6.95	4.99	24.00	
28.03.26	2.18	507.59	505.32		2.277	44.3	40.1	6.82	4.86	24.00	
29.03.26	2.14	507.66	505.42		2.238	44.2	40.1	6.76	4.80	24.00	
30.03.26	2.10	504.50	502.35		2.156	44.2	40.1	6.83	4.89	24.00	
31.03.26	2.17	499.60	497.42		2.180	45.0	40.7	6.87	4.96	24.00	
Итого:	95.30	15624.45	15556.52		67.94	50.8	44.7	6.91	4.96	744.00	

	Q [Гкал]	Gпод [т]	Gобр [т]	Tнар [°C]
31.03.2026 24:00:00	452.33	45203.80	45011.20	2159.99
01.03.2026 00:00:00	357.03	29579.34	29454.69	1415.99
Итого:	95.30	15624.45	15556.52	744.00

Применённые методы

1. Метод анализа предметной области

- Изучены типы ПУ и их диагностические параметры
- Классифицированы неисправности и аномалии
- Проанализирован ручной мониторинг
- Выделены ключевые параметры для диагностики

Итого: 12 диагностических параметров

2. Метод статистического анализа

- Разведочный анализ (EDA): пропуски, выбросы
- Корреляционный анализ и отбор признаков
- Инженерия признаков ($G_diff = |G_1 - G_2|$)
- Анализ распределения нулей и выбросов по классам

Итого: Сформирован набор из 7 признаков

3. Методы машинного обучения

- Обучены и сравнены 4 модели классификации
- Оптимизация гиперпараметров (Optuna)
- Оптимизация порога принятия решений
- Выбрана модель с максимальной полнотой обнаружения

Итого: Random Forest с порогом 0.2098

4. Инженерный и экономический метод

- Клиент-серверная архитектура (FastAPI + Streamlit)
- Извлечение данных из PDF и Excel-отчёты
- Тестирование на исторических данных (900 ведомостей)
- Оценка экономической эффективности внедрения

Итого: Сервис разработан, эффективность подтверждена

Программные инструменты



Python

Язык разработки (бэкенд, ML, анализ данных)
python.org



FastAPI

Фреймворк для бэкенда (REST API)
fastapi.tiangolo.com



Streamlit

Фреймворк для фронтенда (веб-интерфейс)
streamlit.io



scikit-learn

Библиотека для ML (Random Forest, метрики)
scikit-learn.org



Optuna

Библиотека для оптимизации гиперпараметров
optuna.org



pandas

Библиотека для обработки и анализа данных
pandas.pydata.org



NumPy

Библиотека для численных вычислений и работы с массивами
numpy.org



Matplotlib

Библиотека для визуализации данных (графики, диаграммы)
matplotlib.org



pdfplumber

Библиотека для извлечения данных из PDF
github.com/jsvine/pdfplumber



openpyxl

Библиотека для создания Excel-отчётов
openpyxl.readthedocs.io

Разведочный анализ данных (EDA)

От исходных данных к финальному набору признаков

1. Данные и структура

- 10 000 суточных записей-наблюдений
- 5000 исправных / 5000 неисправных
- Разметка выполнена экспертами на основе актов расследования, журналов дефектов и технологических заключений
- Первичный набор: 13 признаков (11 числовых)
- Названия приведены к единому латинизированному формату

2. Пропуски и нули

- Исключены из-за критической доли пропусков:
error (100%), G1_minus (58.24%), G1_plus (48.65%), P1 (3.33%), P2 (2.25%)
- Оставшиеся пропуски заменены на **0** (диагностический признак)
- Нули в Q, G1, G2, T1, T2, Time → **100%** у неисправных
- Нули в G_diff → **96.5%** у неисправных

3. Инженерия признаков

- Создан признак дисбаланса расходов:
 $G_diff = |G_1 - G_2|$
- Рассмотрены два варианта: модуль разности и разность со знаком
- Выбран модуль разности, так как он создаёт монотонную зависимость с целевой переменной

4. Отбор и выбросы

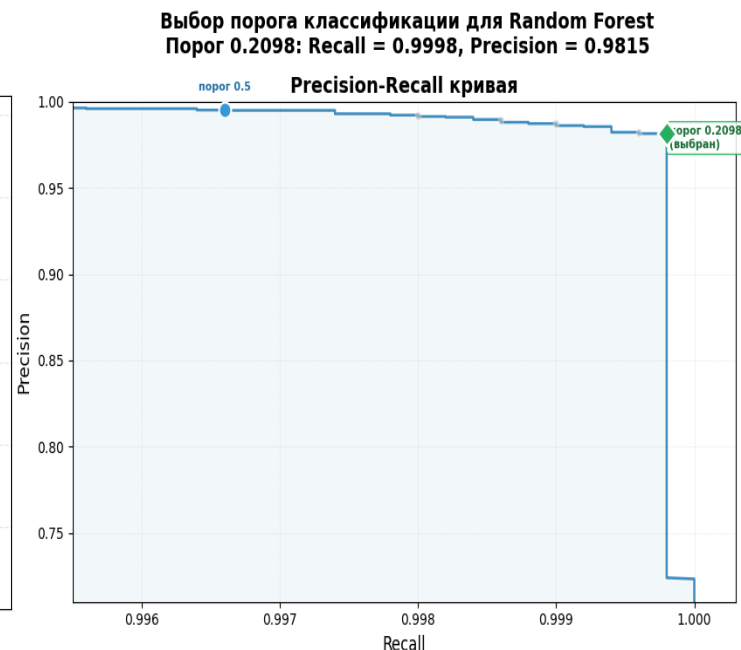
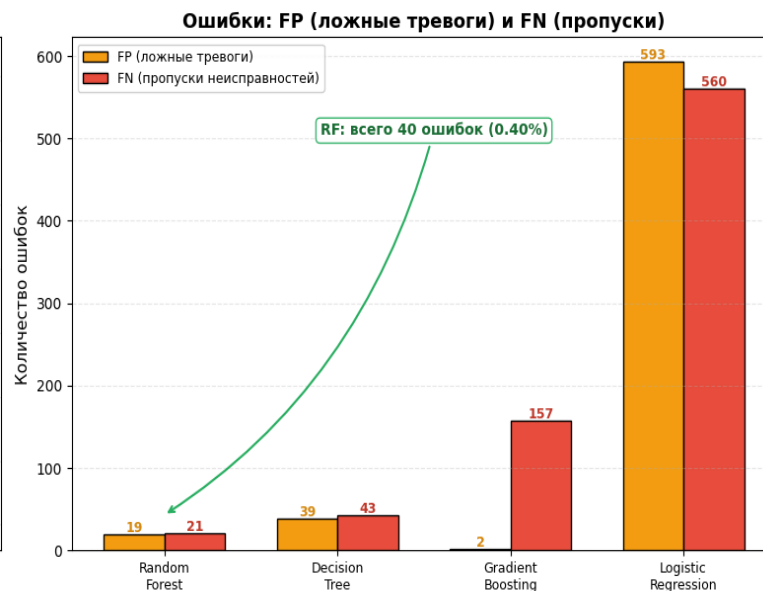
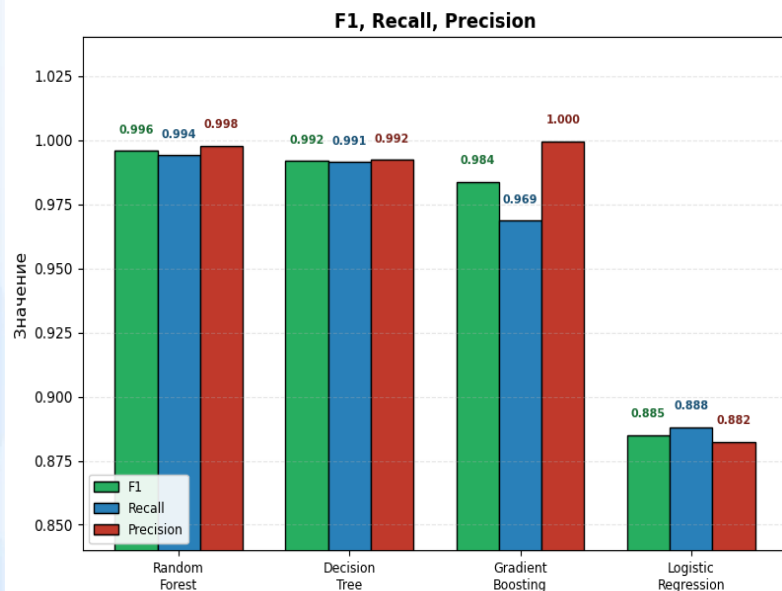
- Проведён корреляционный анализ, выявлена мультиколлинеарность Q, G1, G2 (0.99–1.00)
- Признак G_diff ортогонален остальным, что подтверждает его уникальный вклад
- Финальный набор (7 признаков):
Q, G1, G2, G_diff, T1, T2, Time
- Выбросы в Time → **100%** у неисправных
- G_diff → **99.6%** у неисправных
- T1 → **94.5%**, T2 → **84.1%** у неисправных

Итог EDA: Сформирован финальный набор из **7 признаков**.

Нули и выбросы в ключевых признаках являются **диагностическими маркерами** неисправности.

Выбор и улучшение модели ML

Сравнение моделей (Nested CV + Optuna, 5×3 фолда)
Random Forest — лучший баланс: F1=0.9958, Recall=0.9940, Precision=0.9976



1. Выбор метрик

- Асимметричная цена ошибки: FN → потери, FP → проверка
- Приоритет: **Recall** (полнота)
- Дополнительно: Precision, F1, Accuracy, ROC-AUC

2. Сравнение моделей

- 10000 записей, 7 признаков
- Nested CV + Optuna
- Лучшая: **Random Forest** (40 ошибок)

3. Оптимизация порога

Порог	Recall	Precision	TP	TN	FP	FN
0.5	0.9940	0.9976	4979	4981	19	21
0.2098	0.9996	0.9815	4998	4906	94	2

94 FP — цена за 99.96% обнаружения

4. Итоговые метрики

- Random Forest**, порог 0.2098
- Accuracy: 0.9958
- Precision: 0.9815 (94 FP)
- Recall: 0.9996 (2 FN)
- F1: 0.9906, ROC-AUC: 0.9997

Архитектура сервиса

Клиент-серверная архитектура с модулями визуализации и экспорта



Связи и обратная связь

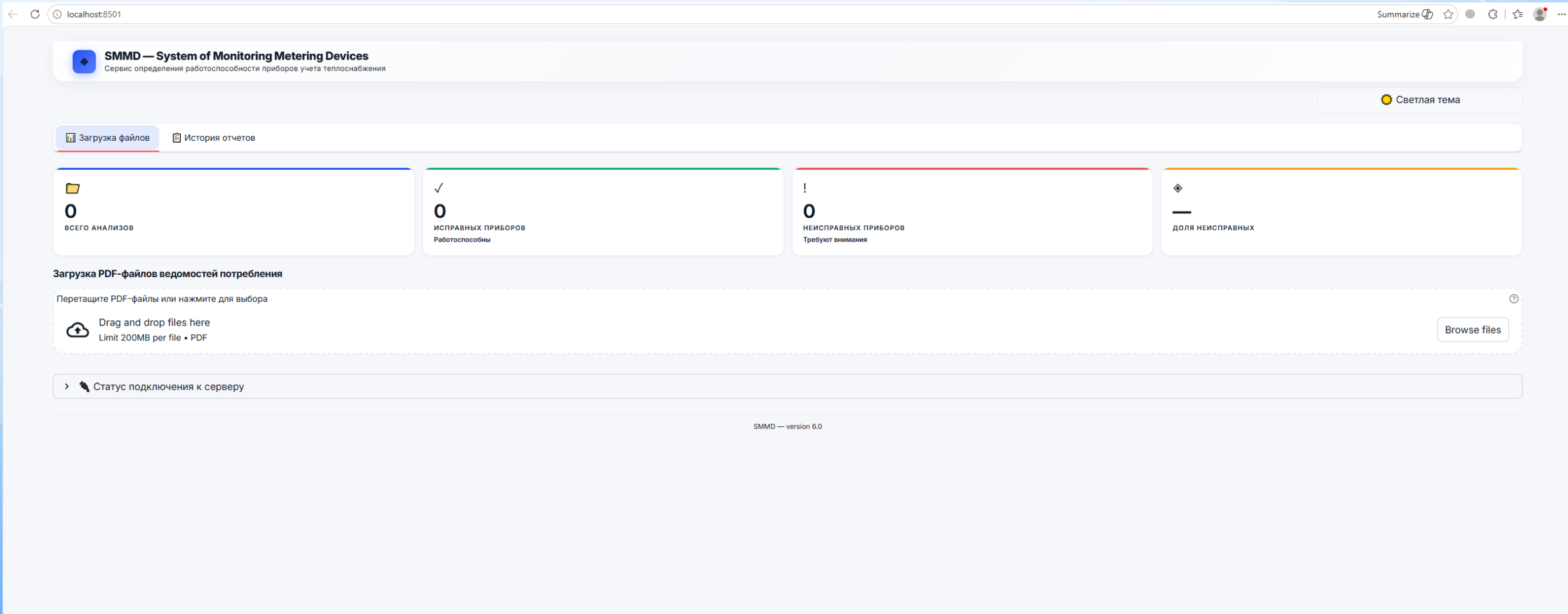
Пользователь ↔ Загрузка PDF ↔ UI-визуализация ↔ Бэкенд-обработка ↔ Excel-отчёт

Итоговая цепочка: Загрузка PDF → Визуализация в UI → Обработка на бэкенде → Возврат результата (JSON + Excel) → Пользователь получает готовое заключение.

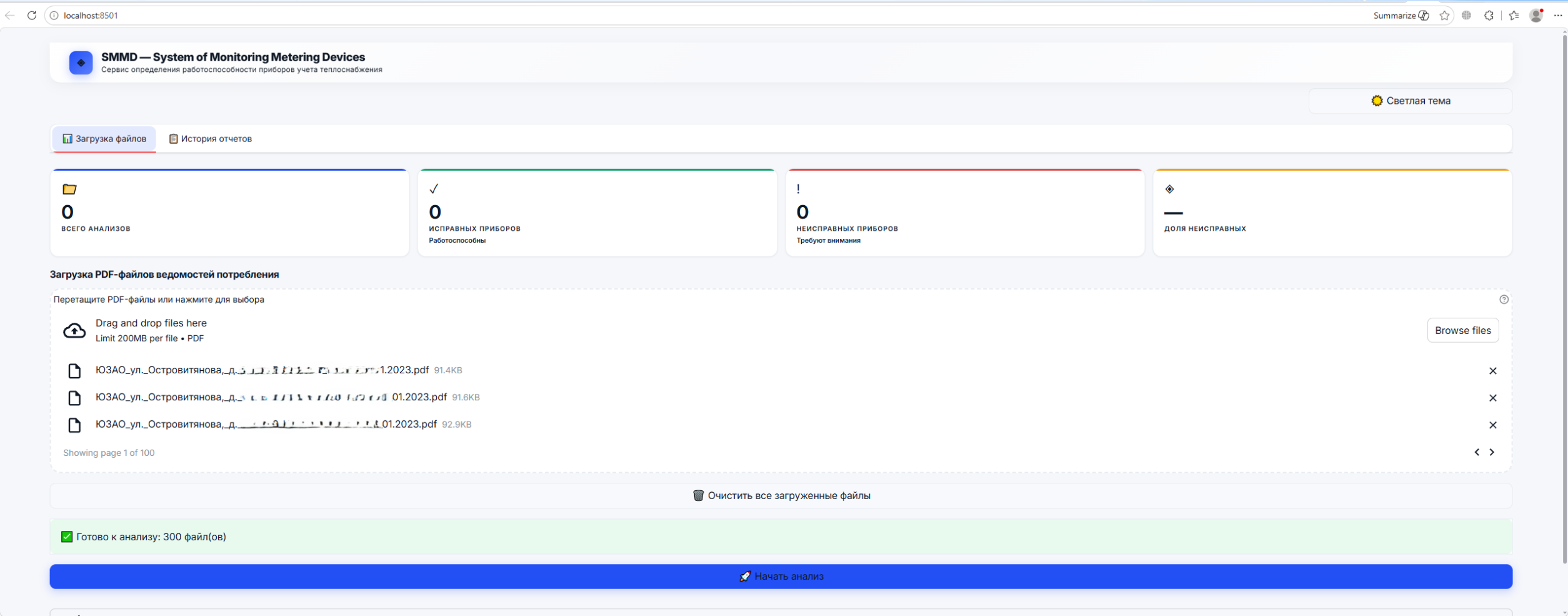
Скриншоты UI сервиса с примерами этапов его работы

Главная страница.

Окна прогресса поэтапных анализов с накоплением результатов. Возможность смены темы (светлая и темная) и проверки статуса



Главная страница.
Пакетная загрузка PDF файлов (ведомостей потребления) для анализа с возможностью быстрой очистки области загрузки



Главная страница.
Визуализация результатов анализа в UI, после анализа загруженных файлов с возможностью скачивания Excel - отчета

🔄

📄

🌐

🔧

🔔

👤

⋮

✅ Анализ завершен

Сводная статистика

ВСЕГО ФАЙЛОВ
300

УСПЕШНО
298

С ОШИБКАМИ
0

ИСПРАВНЫ
276

НЕИСПРАВНЫ
22

Детальные результаты

Файл	Адрес	Статус	Причина	Модель	Вероятность	G ₁ -G ₂ %	Q откл. %	Дата	Пр. даты	T ₁ /T ₂	Время (ч)	Пустые колонки	Проверка
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: - Прибор: - Сер.н	Данные отсутствуют	Показания отсутствуют	ЭЛТЕКО	N/A	N/A	N/A	N/A		0.0/0.0	N/A	Нет данных	Нет данных
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: - Прибор: -	ИСПРАВЕН	Нет аномалий	КМ-5-2	0.0449	1.31	0.34	31.01.2023	⚠️ 1229 дн	87.8/52.7	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: Не указан П	ИСПРАВЕН	Нет аномалий	ТеРосс-ТМ	0.0683	2.59	0.27	31.01.2023	⚠️ 1229 дн	89.5/43.6	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: +7(495)422	ИСПРАВЕН	Нет аномалий	ТеРосс-ТМ	0.0178	2.03	0.20	31.01.2023	⚠️ 1229 дн	90.4/56.9	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: - Прибор: -	ИСПРАВЕН	Нет аномалий	КМ-5-2	0.0814	2.24	0.07	31.01.2023	⚠️ 1229 дн	88.2/60.9	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: +7 (915) 1	ИСПРАВЕН	Нет аномалий	ТеРосс-ТМ	0.0858	0.61	0.07	31.01.2023	⚠️ 1229 дн	79.4/41.3	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Голубинская_ул._д. 11_01.2023.pdf	потребителя: Голубинская ул. Телефон: - Прибор: -	НЕИСПРАВЕН	Обнаружена неисправность (модель)	КМ-5-2	1.0000	0.00	0.00	31.01.2023	⚠️ 1229 дн	45.9/6.8	0.000000	Пустые: Q, G1, G2, G_diff, Time	Дата > 2 сут (+1229дн); Time=0.00 € [23.5, 24.0]
ЮЗАО_Литовский_б-р_д. 11_01.2023.pdf	потребителя: Литовский б-р. Телефон: Не определено	ИСПРАВЕН	Нет аномалий	МКТС	0.0026	0.36	0.44	31.01.2023	⚠️ 1229 дн	87.9/72.5	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Литовский_б-р_д. 11_01.2023.pdf	потребителя: Литовский б-р. Телефон: Неизвестно Пр	ИСПРАВЕН	Нет аномалий	ТеРосс-ТМ	0.0509	2.01	0.09	31.01.2023	⚠️ 1229 дн	86.8/51.9	24.000000	Все колонки заполнены	Дата > 2 сут (+1229дн)
ЮЗАО_Литовский_б-р_д. 11_01.2023.pdf	потребителя: Литовский б-р. Телефон: Неизвестно П	НЕИСПРАВЕН	Обнаружена неисправность (модель)	КМ-5-2	0.9650	0.45	0.08	31.01.2023	⚠️ 1229 дн	86.2/33.1	17.060000	Все колонки заполнены	Дата > 2 сут (+1229дн); Time=17.06 € [23.5, 24.0]

📄 Отчет сохранен: отчет_20260613_144413.xlsx

📄 Скачать отчет (Excel)

💡 Отчет также доступен во вкладке «История отчетов»

Вкладка «История отчетов».

Формирование Excel – отчетов с сохранением файлов в порядке выполнения этапов анализа

SMMД — System of Monitoring Metering Devices

Сервис определения работоспособности приборов учета теплоснабжения

Светлая тема

Загрузка файлов

История отчетов

История отчетов

Отчет от 13.06.2026 14:49

Файл: отчет_20260613_144926.xlsx

Обработано: 300 шт.

Время: 77.0 сек

ВСЕГО

300

УСПЕШНО

298

ИСПРАВНЫ

267

НЕИСПРАВНЫ

31

Загрузить с сервера

Отчеты хранятся 24 часа

Отчет от 13.06.2026 14:47

Файл: отчет_20260613_144708.xlsx

Обработано: 300 шт.

Время: 75.3 сек

ВСЕГО

300

УСПЕШНО

297

ИСПРАВНЫ

262

НЕИСПРАВНЫ

35

Загрузить с сервера

Отчеты хранятся 24 часа

Отчет от 13.06.2026 14:44

Файл: отчет_20260613_144413.xlsx

Обработано: 300 шт.

Время: 81.9 сек

ВСЕГО

300

УСПЕШНО

298

ИСПРАВНЫ

276

НЕИСПРАВНЫ

22

Загрузить с сервера

Отчеты хранятся 24 часа

Пример Excel – отчета

январь — копия для примера отчета - Excel

Поиск

Вход

Поделиться

ФайлГлавнаяВставкаРазметка страницыФормулыДанныеРецензированиеВидСправка

ВырезатьКопироватьВставитьБуфер обмена

Формат по образцу

Шрифт

Выравнивание

Число

Стили

Условное форматирование

Форматировать как таблицу

ОбычныйВыход

НейтральныйВычисление

ПлохойКонтроль...

ХорошийПояснение

ВводПримечание

Ячейки

Редактирование

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Имя файла	Модель ПУ	Серийный номер ПУ	Источник анализ	Статус прибора	Причина	Вероятность неисправности	Кол-во записей	Последняя дата данны	Проверка даты	Дней с последней даты	Q из ведомости	Q расчетно	Отклонени
1	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1_01.2023.pdf	ЭЛТЕКО	6000	Random Forest	Данные отсутствуют	Показания отсутствуют	0,0000	0	N/A	Нет данных	N/A	0,00	0,0000	
2	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1_01.2023.pdf	KM-5-2	4534	Random Forest	ИСПРАВЕН	Нет anomalies	0,0449	31	31.01.2023	Дата (+1229дн)	1229	2,23	2,2212	
3	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1_01.2023.pdf	TePocc-TM	6400	Random Forest	ИСПРАВЕН	Нет anomalies	0,0683	31	31.01.2023	Дата (+1229дн)	1229	2,06	2,0504	
4	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	7000	Random Forest	ИСПРАВЕН	Нет anomalies	0,0178	31	31.01.2023	Дата (+1229дн)	1229	6,77	6,7591	
5	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1 KM-5-2_01.2023.pdf	KM-5-2	3520	Random Forest	ИСПРАВЕН	Нет anomalies	0,0814	31	31.01.2023	Дата (+1229дн)	1229	4,24	4,2322	
6	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	7200	Random Forest	ИСПРАВЕН	Нет anomalies	0,0858	31	31.01.2023	Дата (+1229дн)	1229	3,03	3,0327	
7	ЮЗАО Голубинская ул., д. 1, к. 1, Вент1 KM-5-2_01.2023.pdf	KM-5-2	4492	Random Forest	НЕИСПРАВЕН	Обнаружена неисправность (модель)	1,0000	31	31.01.2023	Дата (+1229дн)	1229	0,00	0,0000	
8	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 MKTC_01.2023.pdf	MKTC	13007	Random Forest	ИСПРАВЕН	Нет anomalies	0,0026	31	31.01.2023	Дата (+1229дн)	1229	2,98	2,9662	
9	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	6300	Random Forest	ИСПРАВЕН	Нет anomalies	0,0509	31	31.01.2023	Дата (+1229дн)	1229	3,45	3,4463	
10	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 KM-5-2_01.2023.pdf	KM-5-2	3261	Random Forest	НЕИСПРАВЕН	Обнаружена неисправность (модель)	0,9650	31	31.01.2023	Дата (+1229дн)	1229	3,87	3,8675	
11	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	6400	Random Forest	ИСПРАВЕН	Нет anomalies	0,0229	31	31.01.2023	Дата (+1229дн)	1229	2,55	2,5505	
12	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	6400	Random Forest	НЕИСПРАВЕН	Обнаружена неисправность (модель)	0,2665	31	31.01.2023	Дата (+1229дн)	1229	1,30	1,3007	
13	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	6400	Random Forest	ИСПРАВЕН	Нет anomalies	0,0372	31	31.01.2023	Дата (+1229дн)	1229	2,40	2,3960	
14	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	7200	Random Forest	ИСПРАВЕН	Нет anomalies	0,0790	30	31.01.2023	Дата (+1229дн)	1229	3,32	3,3245	
15	ЮЗАО Литовский б-р, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	6400	Random Forest	ИСПРАВЕН	Нет anomalies	0,0209	31	31.01.2023	Дата (+1229дн)	1229	2,92	2,9161	
16	ЮЗАО Новоясеневский пр-т, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	6400	Random Forest	НЕИСПРАВЕН	Обнаружена неисправность (модель)	0,2506	30	31.01.2023	Дата (+1229дн)	1229	7,05	7,0546	
17	ЮЗАО пр. Карамзина, д. 1, к. 1, Вент1 TePocc-TM_01.2023.pdf	TePocc-TM	7100	Random Forest	ИСПРАВЕН	Нет anomalies	0,0256	31	31.01.2023	Дата (+1229дн)	1229	2,77	2,7636	
18	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4592	Random Forest	ИСПРАВЕН	Нет anomalies	0,0108	31	31.01.2023	Дата (+1229дн)	1229	7,46	7,4585	
19	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 AT-T-2_01.2023.pdf	AT-T-2	6150	Random Forest	ИСПРАВЕН	Нет anomalies	0,0034	31	31.01.2023	Дата (+1229дн)	1229	5,25	5,2514	
20	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 AT-T-2_01.2023.pdf	AT-T-2	6160	Random Forest	ИСПРАВЕН	Нет anomalies	0,0000	31	31.01.2023	Дата (+1229дн)	1229	6,17	6,1661	
21	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 AT-T-2_01.2023.pdf	AT-T-2	6160	Random Forest	НЕИСПРАВЕН	Обнаружена неисправность (модель)	0,2638	31	31.01.2023	Дата (+1229дн)	1229	7,04	7,0757	
22	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4200	Random Forest	ИСПРАВЕН	Нет anomalies	0,0009	31	31.01.2023	Дата (+1229дн)	1229	5,47	5,4096	
23	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4310	Random Forest	ИСПРАВЕН	Нет anomalies	0,0050	31	31.01.2023	Дата (+1229дн)	1229	8,62	8,6863	
24	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 AT-T-2_01.2023.pdf	AT-T-2	6180	Random Forest	ИСПРАВЕН	Нет anomalies	0,0050	31	31.01.2023	Дата (+1229дн)	1229	5,58	5,5957	
25	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4500	Random Forest	ИСПРАВЕН	Нет anomalies	0,0009	31	31.01.2023	Дата (+1229дн)	1229	6,04	6,0208	
26	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 1) ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4340	Random Forest	ИСПРАВЕН	Нет anomalies	0,0025	31	31.01.2023	Дата (+1229дн)	1229	10,86	10,7598	
27	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 2) ЦО2 KM-5-2_01.2023.pdf	KM-5-2	5800	Random Forest	ИСПРАВЕН	Нет anomalies	0,0000	31	31.01.2023	Дата (+1229дн)	1229	3,98	3,9989	
28	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 3) ЦО1 KM-5-2_01.2023.pdf	KM-5-2	3290	Random Forest	ИСПРАВЕН	Нет anomalies	0,0175	31	31.01.2023	Дата (+1229дн)	1229	0,27	0,2677	
29	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 TePocc-TM_01.2023.pdf	TePocc-TM	5000	Random Forest	ИСПРАВЕН	Нет anomalies	0,0025	31	31.01.2023	Дата (+1229дн)	1229	2,84	2,8302	
30	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 1) ЦО1 AT-T-2_01.2023.pdf	AT-T-2	6120	Random Forest	ИСПРАВЕН	Нет anomalies	0,0034	31	31.01.2023	Дата (+1229дн)	1229	4,00	4,0038	
31	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 2) ЦО2 AT-T-2_01.2023.pdf	AT-T-2	6120	Random Forest	ИСПРАВЕН	Нет anomalies	0,0100	31	31.01.2023	Дата (+1229дн)	1229	5,87	5,9257	
32	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4630	Random Forest	ИСПРАВЕН	Нет anomalies	0,0006	31	31.01.2023	Дата (+1229дн)	1229	2,91	2,9360	
33	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4550	Random Forest	ИСПРАВЕН	Нет anomalies	0,0027	31	31.01.2023	Дата (+1229дн)	1229	3,10	3,0746	
34	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 1) ЦО1 KM-5-2_01.2023.pdf	AT-T-2	6130	Random Forest	ИСПРАВЕН	Нет anomalies	0,0034	31	31.01.2023	Дата (+1229дн)	1229	3,65	3,6541	
35	ЮЗАО Профсоюзная ул., д. 1, к. 1, (Ввод 2) ЦО2 KM-5-2_01.2023.pdf	AT-T-2	6130	Random Forest	ИСПРАВЕН	Нет anomalies	0,0353	31	31.01.2023	Дата (+1229дн)	1229	1,89	1,8885	
36	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО1 KM-5-2_01.2023.pdf	KM-5-2	4210	Random Forest	ИСПРАВЕН	Нет anomalies	0,1556	31	31.01.2023	Дата (+1229дн)	1229	2,43	2,4093	
37	ЮЗАО Профсоюзная ул., д. 1, к. 1, ЦО2 KM-5-2_01.2023.pdf	KM-5-2	4210	Random Forest	ИСПРАВЕН	Нет anomalies	0,0350	31	31.01.2023	Дата (+1229дн)	1229	20,11	20,0839	

Результаты анализа

Сводная статистика

Готово

100%

Тестирование сервиса

Результаты тестирования на исторических данных (2023 г.)

Методика и результаты по месяцам

- Данные:** 900 ведомостей за январь, март и ноябрь 2023 г. (по 300 файлов на месяц)
- Объекты:** одни и те же адреса во всех месяцах
- Порядок:** загрузка → анализ → сравнение с экспертом
- Цель:** проверка устойчивости к сезонным колебаниям и эффективности на реальных данных

Показатель	Январь	Март	Ноябрь
Всего ведомостей	300	300	300
Данные отсутствуют	2	3	2
Модель: исправные	276	262	267
Модель: неисправные	22	35	31
Модель: не найдены (FN)	0	0	1
Эксперт: исправные	292	287	284
Эксперт: неисправные	5	10	14
Эксперт: не найдены	0	1	0

Модель: 0 пропусков в январе и марте; 1 пропуск в ноябре.
Эксперт: 1 пропуск в марте.

Метрики качества и анализ

Метрика	Январь	Март	Ноябрь	Среднее
Accuracy	0,9433	0,9167	0,9367	0,9322
Precision	0,2917	0,3421	0,4545	0,3628
Recall	1,0000	1,0000	0,9375	0,9792
F1-score	0,4516	0,5098	0,6122	0,5245

Recall — приоритетная метрика.

В январе и марте — 100%; в ноябре — 93,75% (1 пропуск).

Анализ: Низкая точность (Precision = 36,28%) объясняется **перестраховкой модели** на данных с пограничными значениями. Эксперту лучше **перепроверить «сомнительные» показания**, чем пропустить неисправность.

Примечание: **Экспертные проверки** не влияют на вердикт модели, а служат дополнительным инструментом для аналитика.

Дообучение: Предусмотрено **периодическое дообучение** по мере накопления новых данных для повышения точности в «серой зоне».

Итог: Сервис подтвердил **устойчивость к сезонным колебаниям и эффективность на реальных данных** (средний Recall = 97,92%). Результаты обосновывают применимость для внедрения.

Расчёт экономической эффективности сервиса и этапы его развития

Параметр	Значение	Обоснование
Приборов учета	30 000 шт.	Крупная городская теплопоставляющая организация
Ведомостей / мес. на 1 ПУ	10 (1 раз в 3 дня)	Регламент предоставления показаний
Всего ведомостей в месяц	300 000 шт.	—
Доля неисправных	5% (1500 шт./мес.)	Расследование около 5% неисправных от общего числа ПУ
Ручной анализ ведомости	2 минуты	Экспертная оценка
Расследование неисправности	15 минут	Оформление, анализ, направление инженерам
Производительность сервиса	300 ведомостей / 75 сек	Замер времени скрипта
Ручная проверка после модели	10%	Пограничные случаи, аномалии, ошибки распознавания
Анализ ведомостей (доля)	67%	Остальное — заявки, коммуникация, задачи
Коэффициент занятости	0,7	70% чистого рабочего времени
Реальный фонд времени	$160 \times 0,7 = 112$ часов	—
Доступно для анализа	$112 \times 0,67 = 75$ часов	—
Час работы аналитика	1 000 руб.	Оценка рынка труда
Зарплата аналитика / мес.	160 000 руб.	До вычета налогов, с учётом премий
Зарплата специалиста сопровождения	250 000 руб.	Администрирование серверов, БД, API, стриминг

Расчёт численности:
 $300\,000 \times 2 \text{ мин} = 10\,000 \text{ ч};$
 $1500 \times 15 \text{ мин} = 375 \text{ ч};$
 Итого: 10 375 ч/мес.
 Доступно: 75 ч/аналитик.
 $10\,375 / 75 = 138,3 \rightarrow 139 \text{ человек.}$

Расчёт затрат:
 $139 \text{ чел.} \times 160\,000 \text{ Р/мес.}$
 $= 22,24 \text{ млн Р/мес.}$
 $\times 12 \text{ мес.}$
 $= 266,9 \text{ млн Р/год.}$

Этап 0. Ручной режим

- Вручную просматривают 300 000 ведомостей/мес.
- Аналитики: **139 человек**
- Затраты: **266,9 млн Р / год**
- Отсутствует автоматизация и оперативная статистика

Этап 1. Локальное развёртывание

- Сервис на рабочих станциях, PDF вручную
- 90% ведомостей отсеиваются автоматически
- Аналитики: **19 человек**
- Затраты: **36,5 млн Р / год**
- Экономия **230,4 млн Р / год (86,3%)**

Этап 2. API-интеграция

- Сервис на сервере, PDF через API
- Централизованная БД (PostgreSQL), отчётность
- Аналитики: **17 человек + 1 админ БД**
- Затраты: **36,2 млн Р / год**
- Экономия **230,6 млн Р / год (87,1%)**

Этап 3. Полная автоматизация

- Интеграция с телеметрией, анализ в реальном времени
- Автоматические заявки в CRM, дашборды, прогнозирование
- Аналитики: **6 человек + DevOps + инженер данных**
- Затраты: **19,3 млн Р / год**
- Экономия **247,6 млн Р / год (94,2%)**

Итог: Переход от ручного режима к полной автоматизации позволяет **сократить штат с 139 до 8 человек** и сэкономить **247,6 млн рублей в год**. Сервис рекомендован к тиражированию в организациях ЖКХ.

Практическая значимость

Снижение трудозатрат — автоматизация сокращает время анализа на **86–94%**. Ручной режим: **139** → локальный: **19 человек**.

Исключение человеческого фактора — единообразное качество диагностики, не зависящее от квалификации специалиста.

Максимальное обнаружение — **99.96%** неисправных приборов (FN = **2**) при **94** ложных срабатываниях.

Предотвращение финансовых потерь — раннее выявление неисправностей защищает финансовые интересы всех сторон.

Широкая область применения — теплоснабжающие организации, УК, администраторы зданий. Масштабируется на **тысячи приборов**.

Доступность — не требует глубоких экспертных знаний: загрузил ведомости — получил готовое заключение.

Направления развития сервиса

Прогнозирование отказов (Predictive Maintenance) — переход от реактивной диагностики к предиктивной («вероятность отказа в ближайшие N дней»).

Расширение типов приборов — ультразвуковые расходомеры, теплосчётчики других производителей, ГВС, вода, газ.

Улучшение интерпретируемости (XAI) — внедрение explainable AI для объяснения решений модели конечным пользователям.

Облачное развёртывание — контейнеризация (Docker, Kubernetes), SLA $\geq 99.9\%$, автоматическое масштабирование.

Мобильное приложение — iOS/Android с push-уведомлениями и просмотром статистики для оперативного реагирования.

Итог: Сервис доказал эффективность на реальных данных и имеет высокий потенциал для масштабирования. Рекомендован к тиражированию в организациях ЖКХ.

Заключение

Ключевые результаты по этапам разработки

АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

12 диагностических параметров

Выявлены ключевые признаки для оценки состояния ПУ

ИНЖЕНЕРИЯ И ОТБОР ПРИЗНАКОВ

7 финальных признаков

Q, G₁, G₂, G_diff, T₁, T₂, **Time**

МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ

Random Forest

Порог **0.2098** · полнота **99.96%** (FN = 2)

РАЗРАБОТКА СЕРВИСА

FastAPI + Streamlit

Клиент-серверная архитектура с визуализацией и экспортом

ТЕСТИРОВАНИЕ НА РЕАЛЬНЫХ ДАННЫХ

97,92%

Средний **Recall** на 900 ведомостях (январь, март, ноябрь)

ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ

139 → 8 человек

Годовая экономия **247,6 млн Р** · сокращение штата на **94,2%**

Итог: Разработан локальный сервис диагностики ПУ на основе ML. Задачи решены в полном объеме.

Список использованных источников

Нормативные документы, учебные издания и техническая документация

1. Постановление Правительства РФ от 18.11.2013 № 1034 «О коммерческом учете тепловой энергии, теплоносителя». — М., 2013.
2. Жерон, О. Прикладное машинное обучение с помощью Scikit Learn, Keras и TensorFlow. — СПб.: БХВ-Петербург, 2020. — 856 с.
3. Маккинни, У. Python и анализ данных. — М.: ДМК Пресс, 2020. — 544 с.
4. Вандер Плас, Дж. Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.
5. Мюллер, А.Ч., Гвидо, С. Введение в машинное обучение с помощью Python. — М.: Альпина Паблишер, 2017. — 400 с.
6. Хасты, Т., Тибширани, Р., Фридман, Дж. Элементы статистического обучения. — М.: ДМК Пресс, 2020. — 768 с.
7. Брюс, П., Брюс, Э. Практическая статистика для специалистов Data Science. — СПб.: БХВ-Петербург, 2021. — 416 с.
8. Документация scikit-learn [Электронный ресурс]
<https://scikit-learn.org/stable/>
9. Документация Optuna [Электронный ресурс]
<https://optuna.org/>
10. Документация FastAPI [Электронный ресурс]
<https://fastapi.tiangolo.com/>
11. Документация Streamlit [Электронный ресурс]
<https://docs.streamlit.io/>
12. Документация pdfplumber [Электронный ресурс]
<https://github.com/jsvine/pdfplumber>