

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА МЕТОДА ПРОГНОЗИРОВАНИЯ ОТКАЗОВ В РАСПРЕДЕЛЁННЫХ СИСТЕМАХ МОНИТОРИНГА НА ОСНОВЕ ТЕЛЕМЕТРИЧЕСКИХ ДАННЫХ И МАШИННОГО ОБУЧЕНИЯ

Котько Е.Н.¹, магистрант, e.kotko@bsuir.by

2026

1. Белорусский государственный университет информатики и радиоэлектроники

Ключевые слова: распределённые системы; телеметрические данные; машинное обучение; прогнозирование отказов; предиктивный мониторинг; Random Forest; ROC-AUC; SLA; time-to-failure.

Аннотация: Представлена экспериментальная оценка метода прогнозирования отказов в распределённых системах мониторинга на основе телеметрических данных и алгоритма Random Forest. Для сопоставления использован эвристический baseline-подход, основанный на фиксированных пороговых значениях ключевых телеметрических метрик. Выполнен сравнительный анализ методов по метрикам качества классификации, интерпретируемости результатов и способности раннего выявления предаварийных состояний. Полученные результаты подтверждают применимость предложенного подхода в задачах предиктивного мониторинга и управления отказами распределённой инфраструктуры.

Введение. Современные распределённые системы генерируют высокообъёмные телеметрические потоки, что требует применения методов раннего выявления отказов. Пороговые методы обеспечивают лишь реактивное обнаружение инцидентов и не позволяют прогнозировать деградацию [1].

В работе рассматривается подход к прогнозированию отказов на основе телеметрических данных и методов машинного обучения, включающий построение признакового пространства на основе временных окон и классификацию состояния системы. При этом сохраняется задача экспериментальной оценки применимости данного подхода для предиктивного мониторинга распределённых систем.

Целью исследования является экспериментальная оценка метода прогнозирования отказов на основе телеметрических данных и машинного обучения в сравнении с эвристическим baseline-подходом. Оценка проводится по метрикам качества классификации (ROC-AUC, accuracy, confusion matrix) и способности модели к раннему выявлению предаварийных состояний системы.

Основная часть. Эксперимент выполнен на телеметрических данных распределённой системы, включающих ключевые метрики состояния инфраструктуры: задержку обработки запросов (latency), количество ошибок (errors), ошибки базы данных (db_errors) и число запросов (requests). Данные представлены в виде временного ряда наблюдений, на основе которого

сформирована бинарная целевая переменная, отражающая состояние системы: нормальное функционирование или предаварийный режим.

Для учёта временной динамики использовано агрегирование метрик по временной оси с формированием признакового пространства для двух подходов: модели машинного обучения на основе Random Forest и эвристического baseline, использующего фиксированные пороговые значения ключевых метрик. Сравнительный анализ выполнен по метрикам качества классификации, устойчивости к изменению телеметрических паттернов и способности раннего выявления деградации [2].

Эксперимент реализован в виде интерактивного дашборда на базе Streamlit [3] с использованием Python, библиотек Matplotlib [4] и scikit-learn [5], что обеспечивает воспроизводимость и интерпретируемость результатов.

На рисунке 1 представлены результаты анализа состояния распределённой системы на основе усреднённых телеметрических метрик и их динамики во времени. Установлено, что средние значения latency составляют 0.823 с, errors – 4.289 событий, DB errors – 2.175 событий, что отражает наличие устойчивой нагрузки и признаков деградации, выражающихся в росте ошибок прикладного уровня (application-level errors: ошибки обработки запросов, сбои сервисной логики, HTTP/API-ошибки) и ошибок базы данных.



Рисунок 1 – Дашборд состояния распределённой системы, деградационной динамики и сравнительной оценки моделей

Временные ряды демонстрируют выраженную деградационную динамику: значения errors систематически доминируют среди метрик, а увеличение latency наблюдается синхронно с ростом нагрузки, что подтверждает наличие коррелированного ухудшения состояния системы и перехода к предаварийному режиму.

Сравнительный анализ моделей показывает, что ROC-AUC ML-модели составляет 0.975, тогда как baseline-подхода – 0.719. Полученный разрыв свидетельствует о существенном превосходстве ML-модели в задаче классификации состояния системы. В отличие от эвристического подхода, основанного на фиксированных порогах, ML-модель обеспечивает более

точное разделение классов и устойчивую интерпретацию многомерных телеметрических зависимостей, что особенно важно при нелинейной деградации инфраструктуры.

Экспериментально подтверждается эффективность применения методов машинного обучения в задачах предиктивного мониторинга распределённых систем.

На рисунке 2 представлены результаты анализа значимости признаков, качества классификации и сравнительного анализа ROC-кривых.



Рисунок 2 – Дашборд интерпретации модели машинного обучения и сравнительного анализа качества классификации

Анализ важности признаков показывает, что наибольший вклад в формирование оценки риска вносит DB_errors, далее следуют errors и requests, тогда как latency имеет минимальное влияние. Это подтверждает доминирующую роль сбоев уровня базы данных и прикладных ошибок в процессе деградации системы [6].

Матрица ошибок демонстрирует корректное разделение состояний «нормальное» и «предаварийное» с низким уровнем ошибок классификации, что подтверждает устойчивость модели к шуму телеметрических данных.

Сравнение ROC-кривых показывает, что базовый эвристический подход лишь незначительно превосходит случайный классификатор, тогда как ML-модель демонстрирует наибольшее отклонение от диагонали и, соответственно, лучшую разделяющую способность. ML-модель обеспечивает более качественное разделение состояний системы и более высокую чувствительность к изменениям телеметрических признаков [7].

На рисунке 3 представлены результаты эксплуатационной оценки предложенной модели в условиях, приближённых к production-сценарию, включая анализ SLA-нарушений, прогноз времени до отказа и корреляционный анализ аномалий телеметрических метрик.



Рисунок 3 – Дашборд эксплуатационной оценки модели: SLA-аналитика, прогноз времени до отказа и корреляционный анализ аномалий

Представлена кривая SLA-нарушений, отражающая моменты превышения порогового значения риска ($risk_ml > 0.7$). Анализ временной динамики показывает, что превышение порогового уровня фиксируется за $t = \Delta$ шагов времени до фактического отказа системы, что подтверждает наличие временного запаса для принятия управляющих решений. Таким образом, модель обеспечивает раннее выявление деградации до наступления критического состояния, что является ключевым требованием для AIOps-систем.

Далее представлена метрика Time-to-Failure, характеризующая оставшийся горизонт работоспособности системы до отказа. Наблюдается обратная зависимость между уровнем риска и временем до отказа: при увеличении $risk_ml$ происходит монотонное снижение Time-to-Failure, что позволяет интерпретировать модель как инструмент количественной оценки приближения инцидента. Данная зависимость подтверждает способность модели формировать интерпретируемый прогноз времени деградации системы в единицах временного шага наблюдения.

Корреляционная матрица показывает, что наибольшая связь с итоговым риском наблюдается у метрик, связанных с нагрузкой и ошибками, при этом *latency* формирует диагональную единичную зависимость (1.0), отражающую самоидентичность признака. Остальные коэффициенты корреляции находятся в низком и умеренном диапазоне, что указывает на распределённый характер влияния признаков на итоговую оценку риска [8].

Совокупный анализ результатов, представленных на рисунке 3, подтверждает, что предложенная модель обеспечивает раннее обнаружение SLA-нарушений, прогнозирование времени до отказа и интерпретируемое выявление причин деградации, что делает её применимой для построения предиктивных систем мониторинга распределённых инфраструктур.

Заключение. В работе выполнена экспериментальная оценка метода прогнозирования отказов в распределённых системах на основе телеметрических данных и методов машинного обучения. Реализованы этапы

формирования временных рядов метрик, построения признакового пространства и обучения модели оценки риска деградации системы с использованием эвристического baseline-подхода для сравнительного анализа.

Результаты показали, что предложенная модель обеспечивает комплексную интерпретацию состояния системы, включая раннее выявление SLA-нарушений, оценку времени до отказа и анализ корреляций телеметрических метрик с уровнем деградации. При различиях в формальных метриках качества (ROC-AUC) модель демонстрирует более высокую чувствительность к динамике изменения состояния системы и позволяет фиксировать предаварийные режимы на ранних стадиях их формирования, что критически важно для задач предиктивного мониторинга.

Практическая значимость работы заключается в применимости разработанного подхода в составе AIOps-систем для мониторинга распределённой инфраструктуры. Предложенная модель обеспечивает повышение надёжности сервисов, сокращение времени реакции на инциденты и переход от реактивного к предиктивному управлению отказами в распределённых системах обработки данных.

Список использованных источников

1. Тед, Я. Изучаем OpenTelemetry. Современный мониторинг систем / Я. Тед, П. Остин. – Астана : Sprint Book, 2025. – 240 с.
2. RandomForestClassifier [Электронный ресурс]. – Режим доступа: <https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>. Дата доступа: 02.05.2026.
3. Streamlit documentation [Электронный ресурс]. – Режим доступа: <https://docs.streamlit.io>. – Дата доступа: 08.05.2026.
4. Matplotlib 3.10.9 documentation [Электронный ресурс]. – Режим доступа: <https://matplotlib.org/stable/users/index.html>. – Дата доступа: 08.05.2026.
5. Рашка, С. Машинное обучение с PyTorch и Scikit-Learn / С. Рашка, Ю. Лю, В. Мирджалили. – Астана : Фолиант, 2024. – 688 с.
6. Новиков, Б. А. Основы технологий баз данных : учебное пособие / Б. А. Новиков, Е. А. Горшкова, Н. Г. Графеева; под ред. Е. В. Рогова. — 2-е изд. — Москва : ДМК Пресс, 2020.
7. Серрано, Л. Грокаем машинное обучение / Л. Серрано, Е. А. Горшкова, Н. Г. Графеева. – СПб. : Питер, 2024. – 512 с.
8. Чарити, М. Обеспечение наблюдаемости ПО / М. Чарити, Ф. Лиз, М. Джордж. – Астана : Sprint Book, 2025. – 352 с.