

Парфенов И. Н.

*МИРЭА — Российский технологический университет (РТУ МИРЭА),
Москва, Россия*

**ПРОГНОЗИРОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ УЗЛОВ СЕТИ
ПЕРЕДАЧИ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ
МАШИННОГО ОБУЧЕНИЯ**

Аннотация. В статье рассматривается задача проактивного управления сетевой инфраструктурой на основе прогнозирования производительности её узлов. Актуальность темы обусловлена ростом объёмов передаваемого трафика, усложнением сетевых топологий и недостаточностью традиционных реактивных методов мониторинга, ориентированных на устранение уже возникших сбоев. Цель работы — разработка системы прогнозирования производительности узлов сети передачи данных на основе методов машинного обучения для повышения эффективности управления сетевыми ресурсами. Предложена информационная модель, описывающая ключевые метрики узлов: загрузку процессора, температуру, использование оперативной памяти и дискового пространства, сетевую задержку и потери пакетов. Описаны система классификации и кодирования разнородных данных, а также этапы предобработки. В качестве математического аппарата применена линейная регрессия с оценкой точности прогноза по средней абсолютной ошибке (MAE). Реализован программный прототип на языке Python с использованием библиотек scikit-learn, pandas, NumPy, Matplotlib и Tkinter; данные хранятся в формате CSV. Прототип строит прогноз на несколько шагов вперёд, классифицирует узлы по уровням риска (низкий, средний, высокий) и формирует текстовые рекомендации по оптимизации. Экспериментальная проверка на наборе синтетических данных показала значения MAE в

диапазоне от 0,03 до 3,1 по различным метрикам, что подтверждает применимость подхода для краткосрочного прогнозирования.

Решение ориентировано на импортозамещение и допускает интеграцию в существующие системы сетевого мониторинга.

Ключевые слова: производительность узлов сети, машинное обучение, линейная регрессия, прогнозирование, сетевые метрики, мониторинг сети, средняя абсолютная ошибка, импортозамещение.

FORECASTING THE PERFORMANCE OF DATA TRANSMISSION NETWORK NODES USING MACHINE LEARNING METHODS

Parfenov I. N., Mitrushkin E. I.

MIREA — Russian Technological University (RTU MIREA), Moscow, Russia

Abstract. The paper addresses the problem of proactive network infrastructure management based on forecasting the performance of its nodes. The relevance of the topic stems from the growth of transmitted traffic, the increasing complexity of network topologies, and the limitations of traditional reactive monitoring methods that are aimed at eliminating failures that have already occurred. The aim of the work is to develop a system for forecasting the performance of data transmission network nodes using machine learning methods in order to improve the efficiency of network resource management. An information model is proposed that describes the key node metrics: CPU load, temperature, RAM and disk usage, network latency, and packet loss. The classification and encoding of heterogeneous data, as well as preprocessing stages, are described. Linear regression is used as the mathematical apparatus, with forecast accuracy assessed by the mean absolute error (MAE). A software prototype was implemented in Python using the scikit-learn, pandas, NumPy, Matplotlib, and Tkinter libraries; data is stored in CSV format. The prototype builds a forecast several steps ahead, classifies nodes by risk level (low, medium, high), and generates textual optimization recommendations. Experimental testing on a synthetic dataset showed MAE values ranging from 0.03

to 3.1 for various metrics, which confirms the applicability of the approach for shortterm forecasting. The solution is focused on import substitution and can be integrated into existing network monitoring systems.

Keywords: network node performance, machine learning, linear regression, forecasting, network metrics, network monitoring, mean absolute error, import substitution.

Введение

Современные сети передачи данных составляют основу цифровой инфраструктуры, обеспечивая работу информационных систем, телекоммуникационных сервисов и облачных платформ. С увеличением объёма передаваемой информации, усложнением топологий и ростом разнообразия типов трафика возрастает нагрузка на отдельные узлы сети — маршрутизаторы, коммутаторы и серверы. В этих условиях задача эффективного управления сетевыми ресурсами с обеспечением требуемого качества обслуживания (QoS) и высокой доступности сервисов приобретает особую актуальность [1].

Одной из ключевых проблем управления сетями остаётся предсказуемость их работы. Традиционные подходы к мониторингу и администрированию построены на реактивной модели — устранении неполадок по мере их возникновения — и не в полной мере отвечают требованиям высоконагруженных систем. Переход к проактивному управлению предполагает заблаговременное выявление потенциальных перегрузок, сбоев и аномалий, что позволяет повысить устойчивость сети, оптимизировать использование ресурсов и снизить эксплуатационные издержки [5].

Эффективным инструментом такого перехода выступают методы машинного обучения (МО), способные анализировать большие объёмы сетевых данных и выявлять скрытые закономерности, недоступные классическим пороговым алгоритмам. Применение МО для прогнозирования

временных рядов сетевых метрик описано в ряде работ [1, 8, 11], а отдельные исследования посвящены прогнозированию трафика и нагрузки в облачных и телекоммуникационных средах [5, 9, 10]. Вместе с тем сохраняется потребность в практико-ориентированных решениях, построенных на открытых и отечественных технологиях и пригодных для интеграции в существующие системы мониторинга в рамках стратегии импортозамещения.

Целью настоящей работы является разработка системы

прогнозирования производительности узлов сети передачи данных на основе методов машинного обучения. Для достижения цели решались задачи: исследование применимых методов МО; построение информационной модели системы; обоснование выбора инструментальных средств; реализация программного прототипа, обеспечивающего сбор, обработку, прогнозирование и визуализацию данных; экспериментальная оценка точности прогнозов; формирование рекомендаций по применению прогнозной информации.

Материалы и методы

Информационная модель и метрики. В центре информационной модели находятся узлы сети — маршрутизаторы, коммутаторы и серверы, через которые проходит информационный поток. Для каждого узла формируется профиль из набора ключевых метрик производительности. Наблюдаемые показатели и пороговые значения, по которым узлу присваивается уровень риска, приведены в таблице 1.

Таблица 1 — Метрики производительности узлов и пороги оценки риска

Метрика	Единица измерения	Средний риск	Высокий риск
Загрузка процессора (CPU Load)	%	> 70	> 85
Температура (Temperature)	°C	> 70	> 80
Потери пакетов (Packet Loss)	%	> 3	> 10

Задержка (Latency)	мс	> 50	> 200
Загрузка диска (Disk Usage)	%	> 75	> 90
Использование ОЗУ (RAM Usage)	%	> 75	> 90

Помимо технических характеристик в модель включаются контекстные параметры (дата, время суток, день недели), тип трафика, используемые протоколы (TCP, UDP, ICMP и др.) и исторические данные о сбоях. Учёт временной динамики метрик — скорости их роста или снижения и повторяющихся паттернов — реализуется средствами анализа временных рядов и позволяет выявлять тренды, связанные с типовыми сценариями нагрузки.

Классификация и кодирование данных. Разнородные входные параметры унифицируются с помощью классификаторов, группирующих метрики по уровню (канальный, сетевой, транспортный), источнику (аппаратные, программные, пользовательская активность) и функциональному назначению. Категориальные признаки (имена протоколов, типы устройств, статусы соединений) преобразуются в числовую форму методами one-hot- и label-кодирования, а для нейросетевых моделей — эмбедингов. Числовые метрики нормализуются для приведения к сопоставимым шкалам.

Применяются минимаксная нормализация и стандартизация по z-оценке, определяемые формулами (1) и (2):

$x' = (x - x_{min}) / (x_{max} - x_{min})$, (1) $z = (x - \mu) / \sigma$, (2) где x — исходное значение метрики; x_{min} , x_{max} — минимум и

максимум по выборке; μ — среднее; σ — среднеквадратическое отклонение.

Математическое обеспечение. Для прогнозирования сетевых метрик рассматривался спектр методов: линейные и статистические модели (линейная регрессия, ARIMA), деревья решений и ансамблевые методы (Random Forest, градиентный бустинг), рекуррентные нейронные сети (LSTM, GRU), а также алгоритмы обнаружения аномалий (Isolation Forest, PCA). В прототипе в качестве базовой модели краткосрочного прогнозирования выбрана линейная

регрессия — за счёт простоты реализации, интерпретируемости и достаточной точности на коротких горизонтах. Прогнозное значение метрики представляется линейной зависимостью от дискретного времени (3):

$\hat{y}(t) = w \cdot t + b$, (3) где $\hat{y}(t)$ — прогнозируемое значение метрики в момент t ; w — коэффициент наклона; b — свободный член. Параметры модели оцениваются методом наименьших квадратов на исторических наблюдениях узла.

Точность прогноза оценивается по средней абсолютной ошибке (MAE) и среднеквадратичной ошибке (RMSE), формулы (4) и (5):

$$MAE = (1/n) \cdot \sum |y_i - \hat{y}_i|, \quad (4)$$

$$RMSE = \sqrt{(1/n) \cdot \sum (y_i - \hat{y}_i)^2}, \quad (5)$$

где y_i — фактическое значение; $\hat{y}_i$ — прогнозное значение; n — число наблюдений в окне проверки. В прототипе качество модели контролируется по MAE на последних пяти точках ряда каждого узла.

Формирование набора данных. В текущей реализации система оперирует архивными и синтетически сгенерированными данными, что позволяет сосредоточиться на отладке аналитических механизмов. Набор формируется генератором, имитирующим работу сетевых узлов: значения метрик обновляются с заданной частотой с плавной случайной вариацией. Температура моделируется как функция загрузки процессора (6):

$$T = T_{base} + (CPU / 100) \cdot T_{range}, \quad (6)$$

где T_{base} — базовая температура; T_{range} — диапазон её изменения; CPU — текущая загрузка процессора. Такая зависимость с добавлением управляемого шума обеспечивает реалистичность тестовых данных.

Интегральная оценка состояния сети выполняется по доле узлов с высоким уровнем риска (7):

$$R = N_{high} / N, \quad (7)$$

где N_{high} — число узлов хотя бы с одной метрикой в зоне высокого риска; N — общее число узлов. При $R > 0,5$ состояние сети считается критическим, при $0,05 < R \leq 0,5$ — «под угрозой», иначе — стабильным.

Инструментальные средства. Прототип реализован на языке Python. Для построения и обучения модели применена библиотека `scikit-learn` (класс `LinearRegression` и метрика `mean_absolute_error`); обработка и подготовка данных выполнены средствами `pandas` и `NumPy`; визуализация временных рядов и прогнозов — `Matplotlib`, интегрированной в графический интерфейс через `backend_tkagg`. Графический интерфейс пользователя построен на стандартной библиотеке `Tkinter` с использованием виджетов `ttk`. Данные о производительности хранятся в формате `CSV`, что упрощает развёртывание прототипа и не требует настройки СУБД [13, 14, 18]. Выбор открытых библиотек и совместимость с отечественными операционными системами (`Astra Linux`, «Альт») соответствуют требованиям импортозамещения [6, 9].

Результаты и обсуждение

Разработанная система построена по модульному принципу и включает модуль сбора данных, блок предобработки, модель прогнозирования, подсистему визуализации и модуль формирования рекомендаций. Поток обработки организован как последовательность «загрузка — очистка — анализ — отображение». На этапе предобработки устраняются пропуски, выполняются нормализация и агрегация для расчёта средних показателей по каждому устройству; подготовленные данные передаются в модель прогнозирования, которая анализирует временные ряды и строит прогноз на заданное число шагов вперёд.

Клиентская часть реализована в виде графического приложения. На главном экране отображается сводная таблица устройств с усреднёнными значениями метрик и автоматически вычисленным уровнем риска; цветовая маркировка строк (зелёный, жёлтый, красный) позволяет быстро выявлять

проблемные узлы, а интегральная оценка состояния сети даёт общее представление об инфраструктуре. По выбору устройства открывается окно с графиками исторической динамики и прогнозных значений по всем метрикам, при этом для каждой метрики выводится значение MAE как индикатор достоверности прогноза. Отдельное представление содержит список узлов под угрозой, а на основе прогнозов формируются текстовые рекомендации по оптимизации (перераспределение нагрузки, проверка охлаждения, настройка маршрутизации, контроль аномального трафика).

Экспериментальная проверка проводилась на синтетическом наборе данных, описывающем десять узлов (пять рабочих станций, маршрутизатор, сервер и три коммутатора) с шагом наблюдения в одну минуту. Для каждого узла строился прогноз на три шага вперёд, после чего вычислялась MAE на контрольном окне. Характерные значения ошибки по узлам и метрикам приведены в таблице 2.

Таблица 2 — Средняя абсолютная ошибка (MAE) прогноза по метрикам для отдельных узлов

Узел	CPU	Темп., °C	Потери, %	Задержка, мс	Диск, %	ОЗУ, %
PC1	2,24	3,08	0,03	0,26	1,19	0,82
PC3	0,98	2,15	0,02	0,21	2,40	1,55
PC5	0,73	1,28	0,02	0,04	0,81	2,82
Router	1,72	1,22	0,23	0,16	1,04	1,18
Switch2	0,85	0,47	0,23	0,09	0,79	1,10

Полученные значения MAE располагаются в диапазоне от 0,03 (потери пакетов) до 3,1 (температура) и для большинства метрик не превышают единиц процента или миллисекунд, что соответствует высокой и средней эффективности прогноза в принятой шкале ($MAE < 5$ — высокая, < 15 — средняя). Наибольшая ошибка наблюдается по температуре и загрузке процессора у узлов с выраженной нелинейной динамикой (например, PC1), что согласуется с ограничениями линейной модели на участках резких изменений. Для метрик с устойчивым поведением (потери пакетов, задержка) ошибка

минимальна. В целом результаты подтверждают применимость линейной регрессии для краткосрочного прогнозирования сетевых метрик и раннего выявления узлов с растущей нагрузкой.

Сопоставление с существующими решениями показывает, что предложенная система занимает нишу лёгких, легко развёртываемых инструментов: в отличие от промышленных платформ мониторинга (Zabbix, Nagios, PRTG), она объединяет визуализацию метрик с прогнозированием и генерацией рекомендаций, не требуя настройки серверной инфраструктуры и проприетарных лицензий. Ограничениями текущей реализации являются использование офлайн-данных и опора на линейную модель; для повышения точности на нелинейных и нестационарных участках перспективен переход к ансамблевым методам и рекуррентным сетям (LSTM, GRU), а также подключение источников данных в реальном времени по протоколам SNMP, NetFlow и через REST API.

Заключение

В работе предложена и реализована система прогнозирования производительности узлов сети передачи данных на основе методов машинного обучения. Разработана информационная модель, описывающая ключевые метрики узлов и контекстные параметры, а также система классификации и кодирования разнородных данных. Обоснован выбор инструментальных средств с ориентацией на открытые и отечественные технологии. Реализован программный прототип, обеспечивающий полный цикл обработки — от подготовки данных до визуализации прогнозов и формирования рекомендаций.

Экспериментальная проверка на синтетических данных показала значения MAE в диапазоне 0,03–3,1 по различным метрикам, что подтверждает пригодность линейной регрессии для краткосрочного прогнозирования и раннего выявления потенциально перегруженных узлов.

Практическая значимость решения состоит в возможности интеграции в существующие средства мониторинга, повышении уровня автоматизации управления сетью и сокращении времени реагирования на инциденты. Дальнейшее развитие связано с переходом к работе в реальном времени, расширением набора моделей за счёт ансамблевых и нейросетевых методов и адаптацией системы под отечественные программно-аппаратные платформы.

Список литературы

1. Аль-Свейти М. Методы машинного обучения для прогнозирования трафика в многоуровневой облачной архитектуре для сервисов автономных транспортных средств // Труды учебных заведений связи. 2022. Т. 8, № 4. С. 89–99.
2. Аршинский В. Л., Провоторов В. А. Применение искусственных нейронных сетей и машинного обучения для прогнозирования успеваемости студентов высших учебных заведений // Инновационные педагогические технологии в образовании. 2024.
3. Ашрапова Л. У., Яхшибоев Р. Э., Атаджанов Ш. Ш. Алгоритмы машинного обучения для прогнозирования финансовых рынков: перспективы и ограничения // Innovations in Science and Technologies. 2024.
4. Иванов К. Внедрение моделей машинного обучения в сетевую безопасность IoT-сетей. Екатеринбург: Уральский федеральный университет, 2024.
5. Киричек Р. В., Кулик В. А. Анализ и прогнозирование трафика современных телекоммуникационных сетей с использованием методов машинного обучения // Вестник Астраханского государственного технического университета. 2023. № 1. С. 75–87.

6. Кузнецов А. В., Иванов К. А. Внедрение моделей машинного обучения в сетевую инфраструктуру для повышения производительности // Новые технологии в радиоэлектронных системах: сб. тр. междунар. конф. 2024. С. 206–209.
7. Рогулин Р. С. Использование методов анализа данных и машинного обучения для прогнозирования и планирования спроса в управлении цепочками поставок // Теоретическая экономика. 2023. № 8. С. 35–50.
8. Смелянский Р. Л., Степанов Е. П. Применение методов с машинным обучением для управления сетевой вычислительной инфраструктурой // Доклады Российской академии наук. Математика, информатика, процессы управления. 2024. Т. 516. С. 103–112.
9. Сунь Цюши. Машинное обучение для оптимизации распределения ресурсов в сетях связи. СПб.: Санкт-Петербургский государственный университет, 2024.
10. Кожевников С. В. Применение линейной регрессии для прогнозирования нагрузки в облачных вычислениях // Научные ведомости Белгородского государственного университета. Серия: Математика. Физика. 2023. № 47 (317). С. 90–95.
11. Назаров И. А., Фёдоров М. В. Сравнительный анализ методов линейной регрессии и нейронных сетей в задачах предсказания параметров трафика сетей связи // Информационно-измерительные и управляющие системы. 2022. Т. 20, № 3. С. 113–119.
12. Курников А. В., Мурашов А. А. Методы прогнозирования параметров компьютерных сетей с использованием временных рядов // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. 9, № 2. С. 54–62.
13. Кузьмин А. А. Применение библиотеки Scikit-learn в задачах прогнозирования параметров ИТ-систем // Прикладные информационные технологии и вычислительная техника. 2022. № 4. С. 45–53.

14. Громов П. И. Реализация графических интерфейсов в Python с использованием Tkinter // Современные информационные технологии и ИТ-образование. 2023. Т. 19, № 2. С. 102–109.
15. Орлов А. А. Методы анализа временных рядов при прогнозировании параметров сетей связи // Информационно-управляющие системы. 2021. № 5. С. 60–67.
16. Захаров В. Н. Применение интеллектуальных алгоритмов в системах мониторинга телекоммуникаций // Информационные системы и технологии. 2022. Т. 20, № 3. С. 88–94.
17. Соколов И. А. Построение прогностических моделей в инженерных ИТсистемах с использованием машинного обучения // Труды вузов. Прикладная информатика и математика. 2023. № 1. С. 39–47.
18. Данилов Ю. В. Использование формата CSV для хранения телеметрических данных в программных системах мониторинга // Вестник информационных технологий. 2023. № 6. С. 73–78.