

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПОТЕРЬ ПАКЕТОВ В WI-FI-СЕТЯХ

^{1,2,3}Ананченко И.В., ²Добровольский Д.К.

¹ФГАОУ ВО «Национальный исследовательский университет ИТМО», Санкт-Петербург;

²ФГБОУ ВО «Санкт-Петербургский государственный технологический институт (технический университет)», Санкт-Петербург, e-mail: ddk.dobr.dobr@mail.ru;

³ФГБОУ ВО Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург;

В статье рассмотрена возможность применения различных алгоритмов машинного обучения для прогнозирования потерь пакетов в беспроводных сетях Wi-Fi на основе параметров трафика и характеристик среды. Для анализа собрана выборка сетевых данных, включающая информацию о параметрах передаваемых пакетов и состоянии окружающей сети. Данные получены с использованием сетевого анализатора и специализированных скриптов, фиксирующих уровень сигнала, отношение сигнал/шум, длину пакета, количество подключённых устройств и загруженность канала. После предварительной обработки данные использовались для построения моделей на основе логистической регрессии, метода опорных векторов, случайного леса, градиентного бустинга и многослойного перцептрона. В процессе анализа сравнивались различные модели машинного обучения по качеству прогнозирования потерь пакетов. Наилучшие результаты показали алгоритмы ансамблевого обучения, такие как градиентный бустинг и случайный лес. Выявлено, что эти модели наиболее эффективно распознают сложные зависимости между параметрами сети и вероятностью возникновения потерь, определены наиболее значимые признаки, влияющие на точность прогноза: количество подключённых клиентов, загруженность канала и соотношение сигнал/шум. Проведённое исследование подтвердило, что методы машинного обучения могут быть успешно применены для предсказания потерь пакетов в Wi-Fi-сетях. Полученные модели и выявленные ключевые признаки могут быть использованы для создания интеллектуальных систем мониторинга и оптимизации работы беспроводной инфраструктуры.

Ключевые слова: машинное обучение, беспроводные сети, потери пакетов, прогнозирование, параметры трафика, качество обслуживания

APPLICATION OF MACHINE LEARNING FOR PREDICTING PACKET LOSS IN WI-FI NETWORKS

^{1,2,3}Anantchenko I.V., ²Dobrovolskii D.K.

¹Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg;

²Saint Petersburg State Technological Institute (Technical University), e-mail: ddk.dobr.dobr@mail.ru;

³Baltic State Technical University "VOENMEH" named after D.F. Ustinov, Saint-Petersburg

The article explores the potential of applying various machine learning algorithms to predict packet loss in Wi-Fi wireless networks based on traffic parameters and environmental characteristics. A dataset of network data was collected, including information on transmitted packet parameters and the state of the surrounding network. The data was obtained using a network analyzer and specialized scripts that recorded signal strength, signal-to-noise ratio, packet length, number of connected devices, and channel utilization. After preprocessing, the data was used to build models based on logistic regression, support vector machines, random forest, gradient boosting, and multilayer perceptron. The performance of different machine learning models in predicting packet loss was compared. The best results were achieved by ensemble learning algorithms, such as gradient boosting and random forest. These models were found to most effectively identify complex relationships between network parameters and the likelihood of packet loss. The most significant features affecting prediction accuracy were identified: the number of connected clients, channel utilization, and signal-to-noise ratio. The study confirmed that machine learning methods can be successfully applied to predict packet loss in Wi-Fi networks. The developed models and identified key features can be utilized to create intelligent systems for monitoring and optimizing wireless infrastructure.

Keywords: machine learning, wireless networks, packet loss, prediction, traffic parameters, quality of service

Введение

В условиях постоянного роста объемов беспроводной передачи данных вопросы обеспечения качества обслуживания (QoS) становятся всё более актуальными. Потери пакетов в Wi-Fi-сетях могут существенно снижать производительность приложений, особенно тех, которые чувствительны к задержкам и

потерям, таких как видеоконференции, VoIP и онлайн-игры [1]. Прогнозирование потерь пакетов позволяет принимать превентивные меры для оптимизации сетевых параметров и улучшения условий эксплуатации [2]. Одним из перспективных подходов к решению рассматриваемой задачи является применение методов машинного обучения (ML) [3].

В отличие от предыдущих исследований, сосредоточенных на применении отдельных алгоритмов (например, Random Forest [2]), работа представляет комплексное сравнение пяти различных моделей машинного обучения, включая ансамблевые и нейросетевые подходы, что позволяет объективно оценить их пригодность к задачам прогнозирования потерь пакетов и определить ключевые параметры среды, влияющие на качество соединения.

Цель исследования – анализ возможности применения различных алгоритмов машинного обучения для прогнозирования потерь пакетов в Wi-Fi-сетях на основе параметров сетевого трафика и характеристик окружающей среды.

Материалы и методы исследования

Для проведения исследования собрана выборка сетевых данных с использованием программного обеспечения Wireshark и специализированных скриптов на Python. Один из таких скриптов выполнял захват сетевых пакетов с помощью библиотеки Pyshark – обёртки над Wireshark. Скрипт запускался на ноутбуке, подключённом к Wi-Fi-сети, и сохранял данные о каждом пакете (время, источник, уровень сигнала, протокол, длина и наличие ошибок) в формате CSV. Помимо этого, второй скрипт с помощью iwconfig и iwlwifi на Linux-платформе снимал информацию о параметрах сигнала и окружения, таких как уровень шума, используемый канал и количество подключённых устройств. Собранные данные затем объединялись в единый датасет с меткой о наличии или отсутствии потерь пакетов. Сбор данных проводился в различных условиях (плотность трафика, количество клиентов, расстояние до точки доступа). В качестве признаков (features) рассматривались следующие параметры: уровень сигнала (RSSI – Received Signal Strength Indicator), длина пакета (байт), отношение сигнал/шум (SNR – Signal-to-Noise Ratio), количество подключённых клиентов и загруженность канала [4].

Предварительная обработка данных включала очистку от пропущенных значений, нормализацию и кодирование категориальных признаков. Для построения моделей прогнозирования использовались логистическая регрессия, метод опорных векторов (SVM), случайный лес (Random Forest), градиентный бустинг (XGBoost) и многослойный перцептрон (MLP). У каждой модели есть свои особенности:

логистическая регрессия – базовая модель классификации, используемая для предсказания вероятности наступления бинарного события;

метод опорных векторов (SVM) – алгоритм строит оптимальную гиперплоскость для разделения классов в пространстве признаков;

случайный лес (Random Forest) – ансамблевый метод, основанный на построении множества деревьев решений и агрегировании их результатов;

градиентный бустинг (XGBoost) – улучшенный ансамблевый алгоритм, обучает модели последовательно, минимизируя ошибку предыдущих;

многослойный перцептрон (MLP) – тип искусственной нейронной сети, состоящий из входного, одного или нескольких скрытых и выходного слоя, способный моделировать сложные нелинейные зависимости [5-7].

Для оценки качества моделей использовались метрики: точность (accuracy), полнота (recall), F1-мера и площадь под кривой ROC (AUC-ROC) [8].

Точность (Accuracy) – доля правильно классифицированных примеров из общего числа наблюдений:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \quad (1)$$

где TP – истинно положительные,

TN – истинно отрицательные,

FP – ложно положительные,

FN – ложно отрицательные.

Полнота (Recall) – способность модели обнаруживать все положительные случаи:

$$Recall = \frac{TP}{(TP + FN)}, \quad (2)$$

F1-мера – гармоническое среднее между точностью (precision) и полнотой:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \quad (3)$$

где

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

AUC-ROC – интегральная метрика, отображающая способность модели различать положительные и отрицательные классы на разных порогах вероятности. Она определяется как площадь под ROC-кривой, которая строится на основе значений True Positive Rate (TPR) и False Positive Rate (FPR) [9]:

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

$$FPR = \frac{FP}{FP + TN}, \quad (6)$$

Значение AUC колеблется от 0 до 1, где 1 соответствует идеальной классификации, а 0.5 – случайному угадыванию (худший результат).

Разделение данных на обучающую и тестовую выборки производилось в пропорции 80:20 с использованием стратифицированной выборки [10].

Результаты исследования и их обсуждение

Каждая модель была протестирована на валидационном наборе данных (20% от всей выборки). Основываясь на результатах тестирования моделей, проведенных нами, была составлена таблица 1 и построены графики AUC-ROC (рисунки 1–5). Значения TPR и FPR для построения AUC-ROC кривых вычислялись программно, а после вычисления были экспортирова-

ны в Excel-таблицу с помощью скрипта на языке Python для дальнейшего использования. Наивысшую точность и полноту продемонстрировали ансамблевые модели случайного леса и градиентного бустинга, с небольшим преимуществом XGBoost, точность которого достигла 92%, F1-мера – 0.85, полнота – 0.85 и AUC-ROC – 0.94. У модели на втором месте – случайного леса – точность 91%, F1-мера – 0.82, преимущество перед MLP в показателе полноты (0.81 у Random Forest против 0.77 у MLP). Модели логистической регрессии и SVM продемонстрировали удовлетворительную точность (89% и 91% соответственно), однако их показатели по полноте оказались ниже, чем у других моделей (0.62 и 0.69 соответственно), что делает их менее предпочтительными в условиях, когда важно минимизировать пропущенные случаи потерь.

Таблица 1

Сравнение эффективности моделей машинного обучения

Модель	Точность	Полнота	F1-мера	AUC-ROC
Логистическая регрессия	0,89	0,62	0,74	0,95
SVM	0,91	0,69	0,80	0,96
Случайный лес	0,91	0,81	0,82	0,95
XGBoost	0,92	0,85	0,85	0,95
MLP	0,91	0,77	0,82	0,95

Источник: составлено авторами.

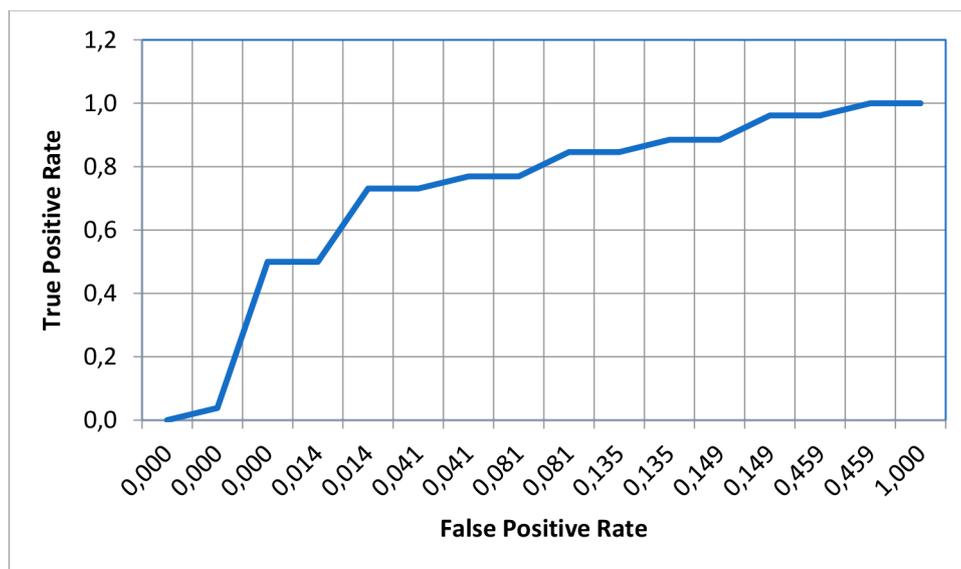


Рис. 1. ROC-кривая для модели логистической регрессии
Источник: составлено авторами

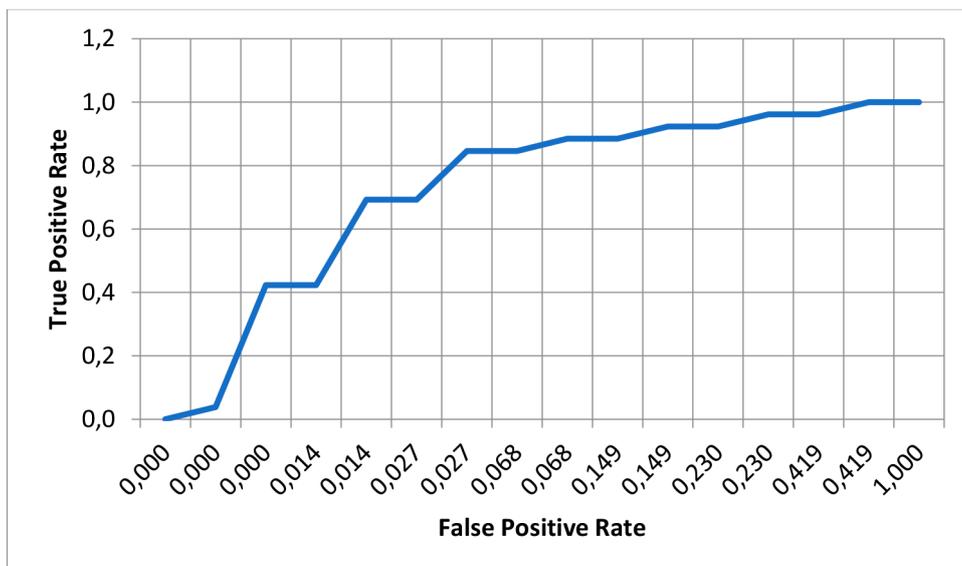


Рис. 2. ROC-кривая для модели SVM
Источник: составлено авторами

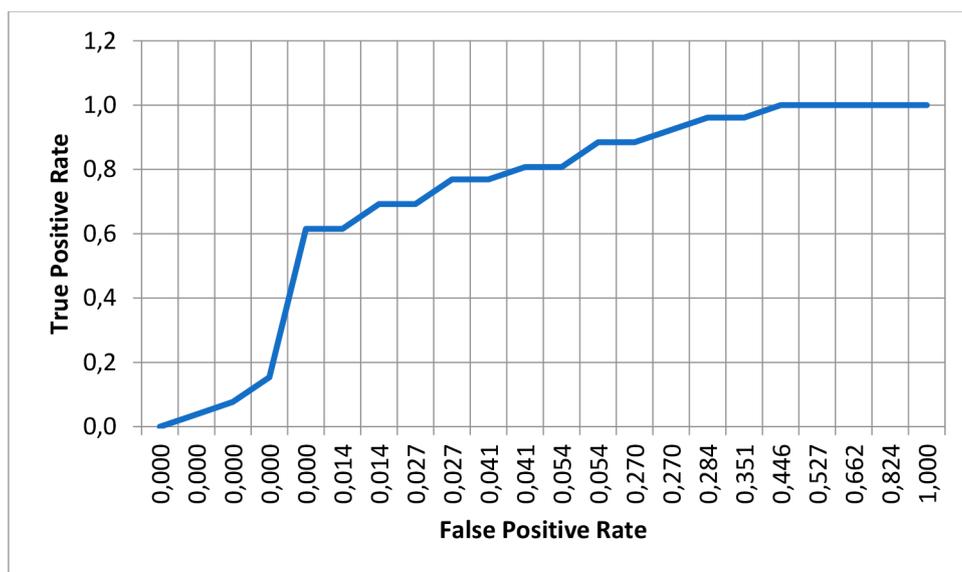


Рис. 3. ROC-кривая для модели случайного леса
Источник: составлено авторами

Результаты показывают, что алгоритмы ансамблевого обучения (Random Forest и XGBoost) лучше справляются с задачей прогнозирования потерь пакетов, скорее всего благодаря тому, что они способны выявлять сложные нелинейные зависимости между признаками [11]. Помимо этого, важность признаков, определённая скриптом

на языке Python с помощью XGBoost, показала, что количество подключённых клиентов, загруженность канала и отношение сигнал/шум (SNR) являются наиболее влиятельными параметрами в случае прогнозирования потери пакетов [12]. Важность признаков (веса признаков), определённая с помощью модели XGBoost, объединена в таблице 2.

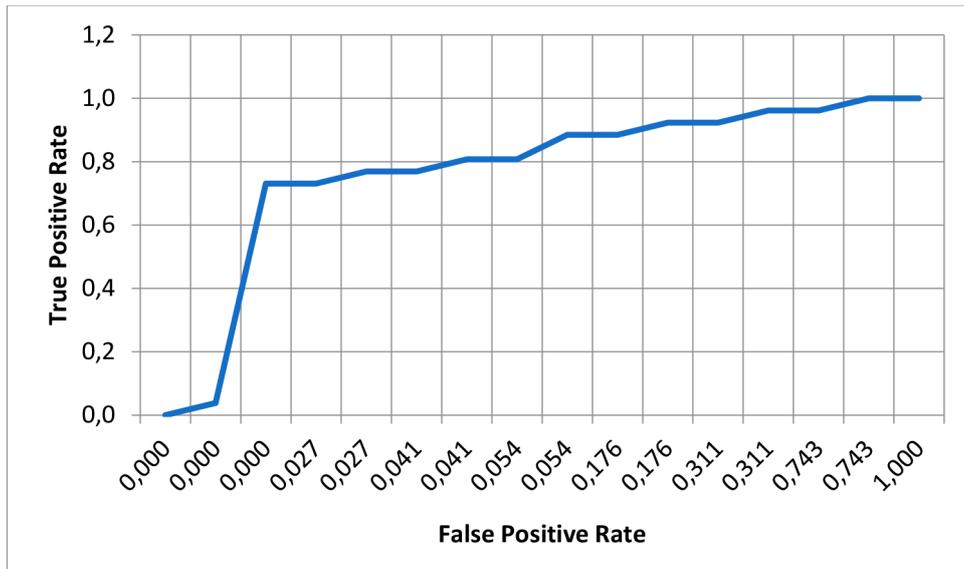


Рис. 4. ROC-кривая для модели XGBoost
Источник: составлено авторами

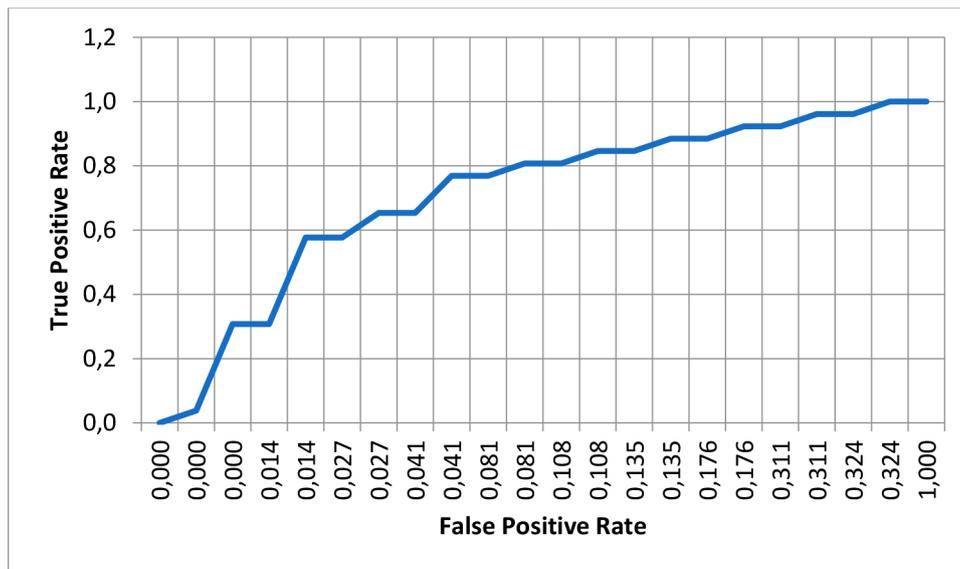


Рис. 5. ROC-кривая для модели MLP
Источник: составлено авторами

Таблица 2

Важность признаков для XGBoost

Признак	Вес признака
Количество подключенных клиентов	0,635
Загруженность канала	0,144
SNR	0,107
Сила сигнала	0,068
Длина пакета	0,045

Источник: составлено авторами.

В отличие от работы Giannakou и др. [2], где рассматривалась лишь одна модель и ограниченный набор признаков, выполненное авторами исследование демонстрирует преимущество ансамблевых и нейросетевых методов при наличии полевых данных различной природы, что способствует расширению понимания применения методов ML к задачам QoS в беспроводных сетях.

Выводы

Проведённое исследование показало, что применение алгоритмов машинного обучения, в частности алгоритмов ансамблевого типа – XGBoost и случайного леса, позволяет с высокой точностью прогнозировать потери пакетов в Wi-Fi-сетях. Эти модели могут быть эффективно интегрированы в системы мониторинга и управления беспроводной инфраструктурой для принятия своевременных корректирующих мер. Выявленные ключевые признаки позволяют сфокусировать внимание на наиболее критичных аспектах при оптимизации параметров сети. В будущем возможно расширение исследования с использованием более обширных и разнообразных исходных данных, а также внедрение моделей в реальное оборудование точек доступа.

Список литературы

1. Bianchi G. Performance analysis of the IEEE 802.11 distributed coordination function // *IEEE Journal on Selected Areas in Communications*. 2000. Vol. 18. No. 3. P. 535–547. DOI: 10.1109/49.840210.
2. Giannakou A., Dwivedi D., Peisert S. A machine learning approach for packet loss prediction in science flows. *Future Generation Computer Systems*. 2020. Vol. 102. P. 190–197. DOI: 10.1016/j.future.2019.08.041.
3. Domingos P. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books, 2015. 352 p. ISBN 978-0-465-06570-7.
4. Wilhelmi F., Góez D., Soto P., Vallés R., Alfaifi M., Algunayah A., Martín-Pérez J., Girletti L., Mohan R., Ramnan K.V., Bellalta B. Machine learning for performance prediction of channel bonding in next-generation IEEE 802.11 WLANs // *ITU Journal on Future and Evolving Technologies*. 2021. Vol. 2. No. 4. DOI: 10.52953/NBGS1213.
5. Bishop C.M. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006. 738 p. (Information Science and Statistics). ISBN 978-0-387-31073-2.
6. Zhang A., Lipton Z.C., Li M., Smola A.J. *Dive into Deep Learning*. Cambridge: Cambridge University Press, 2024. ISBN 978-1-009-38943-3. DOI: 10.1017/9781009389426.
7. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. Cambridge, MA: MIT Press, 2016. 800 p. (Adaptive Computation and Machine Learning series). ISBN 978-0-262-03561-3. URL: <https://www.deeplearningbook.org/> (дата обращения: 24.04.2025).
8. Абдурахимов А.А., Прохoshин А.С. Использование биоинформатических методов и машинного обучения для анализа вегетационных индексов и прогнозирования заболеваний растений // *Наука и инновации: материалы Международной научной конференции молодых учёных* (Ташкент, 15 ноября 2024 г.). Ташкент: Министерство высшего образования, науки и инноваций Республики Узбекистан, Центр передовых технологий, 2024. С. 80–82. DOI: <https://zenodo.org/records/14159997>.
9. Шахбанов З. Метрики в машинном обучении: понимание, применение и интерпретации // shakhbanov.org. [Электронный ресурс]. URL: <https://shakhbanov.org/metrici-v-mashinnom-obuchenii/> (дата обращения: 30.04.2025).
10. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830. DOI: 10.5555/1953048.2078195.
11. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. P. 785–794. DOI: 10.1145/2939672.2939785.
12. Lundberg S.M., Lee S.-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 4765–4774. DOI: 10.48550/arXiv.1705.07874.