

Научная статья
 УДК 004.056.53
 EDN FIEARZ
 DOI 10.17150/2713-1734.2025.7(2).153-164



М.Д. Баженов

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

М.М. Бусько

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

Исследование достоверности обучающих наборов данных анализатора сетевого трафика

Аннотация. В настоящей работе представлены результаты исследования наборов данных нормального и аномального сетевого трафика из общедоступных источников. Целью ставилось определение наиболее качественных данных для глубокого машинного обучения эвристического анализатора способного распознавать вредоносную сетевую активность. В качестве алгоритма машинного обучения выбран метод случайного леса. Проанализировано пять наборов данных. В результате сделано заключение, что набор данных CIC-IDS2017 является наиболее качественным и имеет наиболее высокие значения метрик оценки модели: Precision = 0.95, Recall = 0.94 и F1-score = 0.94.

Кроме этого, были выделены наиболее важные признаки сетевого трафика для классификации нормального и атакующего трафика. В силу того, что признаки в разных наборах данных не совпадают, был составлен унифицированный перечень аналогичных. Наибольший вес в классификации имеют объём переданных данных (total_bytes) и скорость передачи (data_rate). Далее следует длительность соединения (flow_duration).

Результаты, которые получены в настоящей работе, могут быть использованы для тестирования, других алгоритмов машинного обучения и разработки эвристических анализаторов на основе искусственного интеллекта и машинного обучения.

Ключевые слова. Информационная безопасность, анализ сетевого трафика, машинное обучение, набор данных, метод случайного леса.

Информация о статье. Дата поступления: 4 марта 2025 г.; дата принятия к публикации: 11 июня 2025 г.; дата онлайн-размещения: 8 июля 2025 г.

Original article

M.D. Bazhenov

*Baikal State University,
Irkutsk, Russian Federation*

M.M. Busko

*Baikal State University,
Irkutsk, Russian Federation*

Research of the Reliability of Network Traffic Analyzer Training Datasets

Abstract. This paper presents the results of studying datasets of normal and abnormal network traffic from publicly available sources. The goal was to determine

the highest quality data for deep machine learning of a heuristic analyzer capable of recognizing malicious network activity. The random forest method was selected as a machine learning algorithm. Five datasets were analyzed. As a result, it was concluded that the CIC-IDS2017 dataset is of the highest quality and has the highest values of the model evaluation metrics: Precision = 0.95, Recall = 0.94 and F1-score = 0.94. In addition, the most important network traffic features were identified for classifying normal and attack traffic. Due to the fact that the features in different datasets do not coincide, a unified list of similar ones was compiled. The volume of transferred data (`total_bytes`) and the transfer rate (`data_rate`) have the greatest weight in the classification. Next comes the duration of the connection (`flow_duration`).

The results obtained in this work can be used for testing other machine learning algorithms and developing heuristic analyzers based on artificial intelligence and machine learning.

Keywords. Information security, network traffic analysis, machine learning, dataset, random forest method.

Article info. Received 4 March 2025; Accepted 11 June, 2025; Available online 8 July, 2025.

Введение

Анализ сетевого трафика (Network Traffic Analysis, NTA) — это процесс сбора, изучения и интерпретации данных о передаче информации в сети с целью выявления аномалий, проблем с производительностью и вредоносной активности. Традиционные методы такого анализа по большей части основаны на сигнатурном обнаружении, при котором трафик сопоставляется с известными шаблонами вредоносных атак. Однако данный подход малоэффективен против новых или неизвестных угроз, которые ещё не были идентифицированы. Так же очевидным недостатком такого подхода является невозможность анализа зашифрованного трафика, а большой процент данных в корпоративных сетях передается с использованием технологий шифрования VPN. Соответственно, встает необходимость разработки новых эвристических алгоритмов классификации сетевого трафика.

Для построения современных систем анализа, способных выявлять, в том числе атаки нулевого дня рекомендуется применять методы искусственного интеллекта и глубокого обучения [1–3]. Более того эти методы уже реализованы в некоторых коммерческих программах, например, Cisco Stealthwatch¹, PT Network Attack Discovery², Darktrace³ и др. Есть и широко известные открытые проекты доступные на GitHub⁴: DeepPacket⁵, DeepFlow, DeepIntrusion⁶.

¹ <https://www.cisco.com/c/en/us/products/collateral/security/stealthwatch/data-sheet-c78-739398.html>.

² <https://www.ptsecurity.com/ru-ru/products/network-attack-discovery/#page-start>.

³ <https://darktrace.com/news/darktrace-delivers-new-innovations-in-network-detection-and-response-for-enhanced-detection-scalability-and-soc-efficiency-in-the-modern-enterprise>.

⁴ <https://github.com/sdnds-tw/DeepPacket>.

⁵ <https://github.com/caesar0301/deepflow>.

⁶ <https://github.com/deepintrusion/deepintrusion>.

Разработка анализатора сетевого трафика требует выбора адекватного алгоритма машинного обучения, но не менее важным является наличие адекватных данных, на основе которых происходит обучение системы. Для обучения таких систем используют наборы сетевых данных, представляющие собой как нормальный, так и аномальный трафик. Чем достовернее данные, тем точнее система будет диагностировать аномалии и вредоносную активность. Понятно, что многие крупные компании, если и проводят захват реального трафика, то вряд ли поделятся этими данными. Таким образом встает проблема, где брать исходные данные для обучения анализатора.

В свободном доступе имеются обучающие наборы данных, однако гарантии достоверности их никто дать не может. Более того, имеются проблемы, связанные с различием в форматах и разными способами разметки. В настоящей статье проводится исследование датасетов из общедоступных источников. Авторы попытались определить наиболее качественные наборы данных для обучения системы анализа сетевого трафика.

Наборы исследуемых данных и алгоритм обучения

Для обучения системы анализа сетевого трафика требуются наборы данных с нормальным и вредоносным сетевым трафиком. Исследование проводилось на наборах данных из открытых источников, имеющих формат *.csv и схожий состав признаков. Из наиболее известных и опубликованных в открытых источниках удалось получить:

UNSW-NB15 — набор данных для обнаружения сетевых вторжений Австралийского центра кибербезопасности (ACCS)⁷. Это гибридный набор реальных нормальных действий в сети и синтезированные действия атак;

CTU-13 — набор данных о трафике ботнет-сетей, который был собран в Чешском техническом университете (CTU)⁸;

CIDDS-001 — набор данных для исследований обнаружения вторжений в сети опубликованный Кобургским университетом прикладных наук и искусств (Германия)⁹;

NSL-KDD — является усовершенствованной версией набора данных KDD cup99, использовавшегося в третьем международном конкурсе инструментов для интеллектуального анализа данных¹⁰;

CIC-IDS2017 — это набор данных для оценки обнаружения вторжений, подготовленный Канадским институтом кибербезопасности (CIC)¹¹.

⁷ <https://ieee-dataport.org/documents/unswnb15-dataset#files>.

⁸ <https://mcfp.felk.cvut.cz/publicDatasets/CTU-13-Dataset/>.

⁹ <https://www.hscoburg.de/fileadmin/hscoburg/WISENT-CIDDS-001.zip>.

¹⁰ <https://www.unb.ca/cic/datasets/nsl.html>.

¹¹ <https://www.unb.ca/cic/datasets/ids2017.html>.

Задача обоснования выбора алгоритма машинного обучения в настоящей работе не ставилась, потому был взят метод случайного леса (Random Forest), так как он обеспечивает высокую точность и хорошую интерпретируемость результатов. Это подтверждается рядом публикаций в которых дается высокая оценка этого метода применительно к мониторингу сетевого трафика [4–6].

К сожалению, исследуемые наборы данных включают разные составы признаков пространств. Соответственно, обучение алгоритма и проверка его работы проводились на тренировочных и тестовых выборках, принадлежащих одному набору данных.

Инструментарий исследования

Метод случайного леса (Random Forest) достаточно универсальный алгоритм машинного обучения, применимый для широкого круга задач классификации, регрессии и кластеризации. Метод основан на использовании большого ансамбля решающих деревьев в котором каждое дерево дает не высокое качество классификации, а совокупность большого количества их значительно улучшает результат [7]. Для реализации метода использовались возможности языка Python и в частности библиотека Scikit-Learn, предоставляющая унифицированный интерфейс к широкому спектру алгоритмов. Её модульная архитектура и обширная документация существенно упрощают процесс экспериментов с различными моделями и их ансамблями в контексте анализа сетевого трафика. Библиотека поддерживает различные методы классификации, такие как логистическая регрессия и k-ближайшие соседи, метод опорных векторов, наивный байесовский классификатор, дерево принятия решений, а также ансамбль методов, такие как метод случайного леса, алгоритм AdaBoost и градиентный бустинг [8; 9].

Перед обучением модели важно правильно подготовить данные, включая обработку пропусков, нормализацию и разделение на тренировочные и тестовые выборки (рис. 1).

Затем следует этап обучения модели. В нашем случае модель случайного леса обучалась на тренировочных данных, используя 100 деревьев в ансамбле (рис. 2)

Затем обученная модель применялась к тестовой выборке. Здесь для задачи классификации решение осуществлялось голосованием по большинству с использованием метода RandomForestClassifier библиотеки Scikit-Learn. Далее оценивались результаты точности модели (Accuracy), которая отражает долю правильно классифицированных данных среди всех. Также были задействованы такие метрики оценки качества модели как точность предсказания (Precision), полнота (Recall) и гармоническое среднее между точностью и полнотой (F1-score).

```

[6] # Кодирование категориальных признаков
label_encoders = {}
for col in df.select_dtypes(include=['object']).columns:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le

[7] # Разделение на признаки и целевую переменную
X = df.drop(columns=['label']) # 'label' - целевой признак
y = df['label']

[8] # Разделение на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[9] # Проверяем, что все признаки числовые
X_train = X_train.apply(pd.to_numeric, errors='coerce').fillna(0)
X_test = X_test.apply(pd.to_numeric, errors='coerce').fillna(0)

[10] # Масштабирование данных
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

```

Рис. 1. Фрагмент подготовки исследуемых данных

```

✓ 1 мин. [17] from sklearn.ensemble import RandomForestClassifier
        from sklearn.metrics import classification_report, accuracy_score

        # Инициализация и обучение модели
        rf = RandomForestClassifier(n_estimators=100, random_state=42)
        rf.fit(X_train, y_train)

        # Прогнозирование на тестовой выборке
        y_pred = rf.predict(X_test)

        # Оценка результатов
        accuracy = accuracy_score(y_test, y_pred)
        print(f"Accuracy: {accuracy:.4f}")
        print(classification_report(y_test, y_pred))


```

Рис. 2. Обучение модели

Исследование наборов данных

После обучения модели на наборе данных UNSW-NB15 и применения её к тестовой выборке были получены следующие результаты (рис. 3).

Затем обучение и тестирование модели производилось на других наборах данных. Результаты оценки качества этих моделей представлены на рис. 4–7.



Результаты для UNSW-NB15:

Accuracy: 0.8670

Precision: 0.8812

Recall: 0.8670

F1-score: 0.8640

Подробный отчёт по классам:

	precision	recall	f1-score	support
0	0.96	0.74	0.83	37000
1	0.82	0.97	0.89	45332
accuracy			0.87	82332
macro avg	0.89	0.85	0.86	82332
weighted avg	0.88	0.87	0.86	82332

Рис. 3. Результаты обучения модели на наборе данных UNSW-NB15

Результаты модели Random Forest для CTU-13 (с взвешиванием классов, подвыборка 50,000 записей):

Accuracy: 0.9766

Precision: 0.7654


Recall: 0.4164

F1-score: 0.5394

Полный отчёт классификации:

	precision	recall	f1-score	support
Normal	0.98	1.00	0.99	9671
Attack	0.77	0.42	0.54	329
accuracy			0.98	10000
macro avg	0.87	0.71	0.76	10000
weighted avg	0.97	0.98	0.97	10000

Рис. 4. Результаты обучения модели на наборе данных CTU-13



Результаты для CIDD5-001:


Accuracy: 0.9995

Precision: 0.9104

Recall: 0.8883

F1-score: 0.8992

Рис. 5. Результаты обучения модели на наборе данных CIDD5-001



Результаты для NSL-KDD:

Accuracy: 0.7813

Precision: 0.9685

Recall: 0.6366

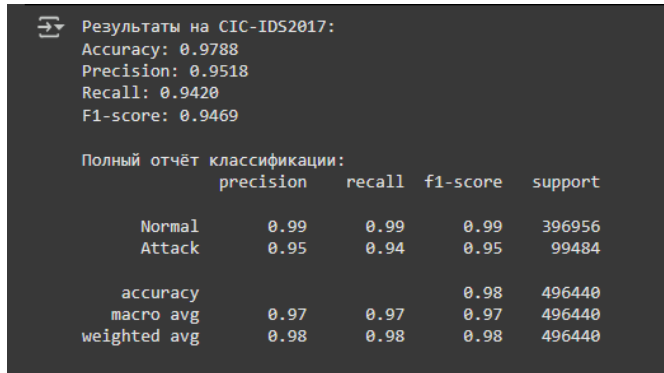
F1-score: 0.7682

Полный отчет классификации:

	precision	recall	f1-score	support
Normal	0.67	0.97	0.79	9711
Attack	0.97	0.64	0.77	12833
accuracy			0.78	22544
macro avg	0.82	0.80	0.78	22544
weighted avg	0.84	0.78	0.78	22544

Рис. 6. Результаты обучения модели на наборе данных NSL-KDD

158



Результаты на CIC-IDS2017:

Accuracy: 0.9788
Precision: 0.9518
Recall: 0.9420
F1-score: 0.9469

Полный отчёт классификации:

	precision	recall	f1-score	support
Normal	0.99	0.99	0.99	396956
Attack	0.95	0.94	0.95	99484
accuracy			0.98	496440
macro avg	0.97	0.97	0.97	496440
weighted avg	0.98	0.98	0.98	496440

Рис. 7. Результаты обучения модели на наборе данных CIC-IDS2017

Результаты значений метрик качества моделей сведены в табл. 1.

Таблица 1

Результаты оценки качества обнаружения атак

Набор данных	Accuracy	Precision	Recall	F1-score
UNSW-NB15	0.86	0.88	0.86	0.86
CTU-13	0.97	0.76	0.41	0.53
CIDDS-001	0.99	0.91	0.88	0.89
NSL-KDD	0.78	0.96	0.63	0.76
CIC-IDS2017	0.97	0.95	0.94	0.94

Как видим из табл. 1, наивысшую оценку точности модели (Accuracy) 0,99 получило обучение на наборе данных CIDDS-001. Однако эта метрика становится практически бесполезной в задачах с неравными классами [10]. Понятно, что в задачах выявления вредоносного трафика, нормальный трафик будет преобладать. Наиболее объективную картину дают метрики Precision и Recall. При этом Precision показывает долю пакетов, названных классификатором вредоносными и при этом действительно являющимися вредоносными. Recall же показывает, какую долю вредоносных пакетов из всех вредоносных нашел алгоритм. Согласно этим оценкам наилучшим набором данных для обучения системы анализа сетевого трафика является CIC-IDS2017.

Наконец, есть агрегированный критерий качества, объединяющий Precision и Recall — это среднее гармоническое F1-score. Метрика F1-score имеет максимальное значение при точности предсказания и полноте равными единице, и близка к нулю, если хотя бы одно из составляющих близко нули. F1-score так же показывает наилучший результат для набора данных CIC-IDS2017.

Оценка важности признаков

Как говорилось ранее, признаковый состав разных наборов данных не совпадает, поэтому оценка важности признаков и их анализ проводились для каждого набора данных в отдельности. Результаты оценки представлены на рис. 8–12.

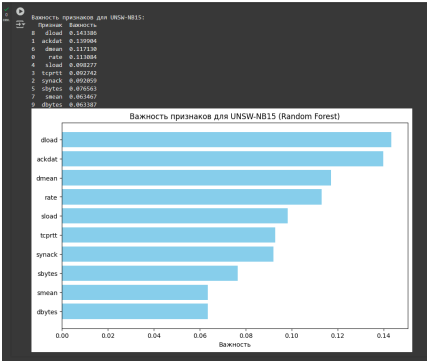


Рис. 8. Важность признаков для UNSW-NB15

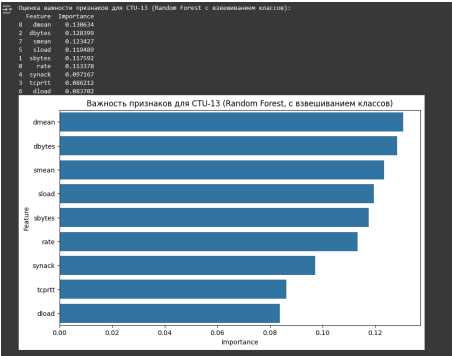


Рис. 9. Важность признаков для STU-13

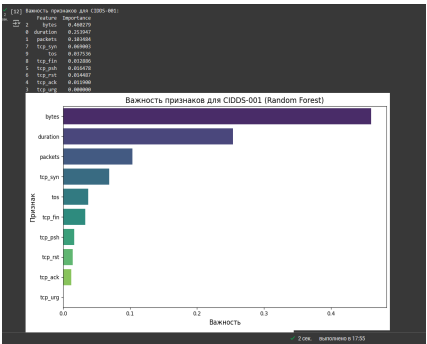


Рис. 10. Важность признаков для CIDD5-001

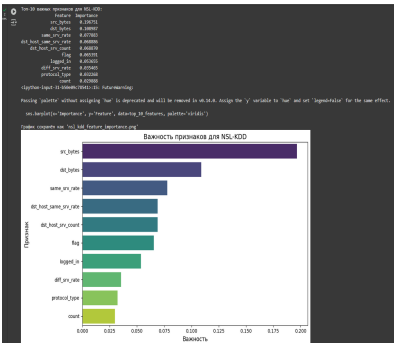


Рис. 11. Важность признаков для NSL-KDD

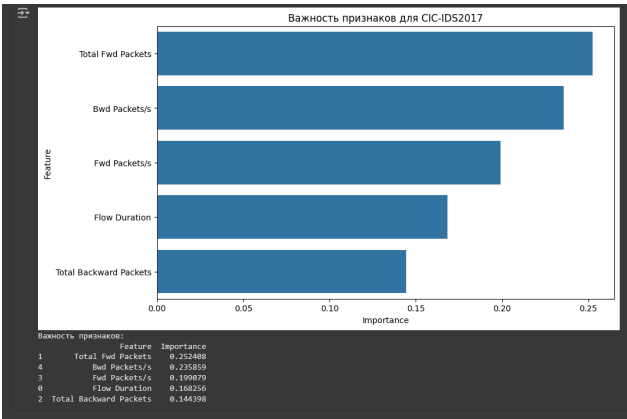


Рис. 12. Важность признаков для CIC-IDS2017

Полученные результаты были проанализированы и выделены общие или аналогичные признаки характерные для нескольких наборов данных и имеющие значимый вес при классификации. В результате анализа был составлен унифицированный перечень, который представлен в табл. 2.

Таблица 2

Важность признаков

Унифицированный признак	Описание	Важность в датасетах
flow_duration	Длительность соединения	CIDDS-001 (0.254), CIC-IDS2017 (0.168)
total_bytes	Общий объем переданных байтов	UNSW-NB15 (sbytes: 0.077, dbytes: 0.063), CTU-13 (sbytes: 0.118, dbytes: 0.128), NSL-KDD (src_bytes: 0.197, dst_bytes: 0.109), CIDDS-001 (bytes: 0.460)
total_packets	Общее количество пакетов	CIC-IDS2017 (Total Fwd Packets: 0.252, Total Backward Packets: 0.144), CIDDS-001 (packets: 0.103)
data_rate	Скорость передачи данных	UNSW-NB15 (rate: 0.113, sload: 0.098, dload: 0.143), CTU-13 (rate: 0.113, sload: 0.119), CIC-IDS2017 (Fwd Packets/s: 0.199, Bwd Packets/s: 0.235)
tcp_handshake	Характеристики TCP-рукопожатия	UNSW-NB15 (synack: 0.092, ackdat: 0.139), CTU-13 (synack: 0.097), CIDDS-001 (tcp_syn: 0.069)
tcprtt	Время отклика TCP	UNSW-NB15 (tcprtt: 0.093), CTU-13 (tcprtt: 0.086)
packet_stats	Статистики размеров пакетов	UNSW-NB15 (smean: 0.063, dmean: 0.117), CTU-13 (smean: 0.123, dmean: 0.130)

На основании табл. 2 можно отметить, что объем переданных данных (total_bytes) и скорость передачи (data_rate) демонстрируют стабильно высокую значимость во всех наборах. Это указывает на их универсальность для классификации нормального и атакующего трафика. Длительность соединения (flow_duration) выделяется как ключевой временной признак в более новых датасетах CIDDS-001 и CIC-IDS2017, тогда как характеристики TCP-рукопожатия (tcp_handshake) и статистики пакетов (packet_stats) имеют контекстно-зависимое влияние, проявляясь в основном в UNSW-NB15 и CTU-13. Такой подход позволяет выявить общие закономерности и различия между наборами данных, облегчая выбор оптимальных признаков для задач обнаружения атак.

Заключение

Наибольшую точность (Ассигу) модели показало обучение на наборе данных CIDDS-001 и составило 99 % (табл. 1), но эта оценка не является объективной в задачах, имеющих неравные

классы. Оценки Precision, Recall и F1-score для атакующего трафика дали максимальные значения, 0.95, 0.94, 0.94 соответственно для модели, обученной на наборе данных CIC-IDS2017, что указывает на высокую эффективность обнаружения атак. Это подтверждает, что данная модель имеет хорошую способность классифицировать как нормальный, так и атакующий трафик. Метрики показывают, что модель хорошо справляется с задачей классификации, демонстрируя высокую точность и сбалансированность между классами.

Таким образом, можно констатировать, что набор данных CIC-IDS2017 является наиболее качественным для обучения системы анализа сетевого трафика.

Результаты анализа признакового пространства показали, что наиболее значимыми характеристиками для модели являются признаки объём переданных данных (total_bytes) и скорость передачи (data_rate), а также длительность соединения (flow_duration). Это объясняется тем, что данные признаки отражают критические аспекты сетевого трафика.

Список использованной литературы

1. Иванов С.О. Методика создания и обучения искусственной нейронной сети для решения задачи распознавания аномалий сетевого трафика / С.О. Иванов. — DOI 10.17587/it.30.32-41. — EDN NRGZNE // Информационные технологии. — 2024. — Т. 30, № 1. — С. 32–41.
2. Поздняк И.С. Модели обнаружения атак с использованием методов машинного обучения / И.С. Поздняк, И.С. Макаров. — DOI 10.18137/RNU.V9187.24.01.P.99. — EDN MNMSYZ // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. — 2024. — № 1. — С. 99–109.
3. Анализ и отбор значимых характеристик сетевого трафика для использования в машинном обучении / Е.А. Дмитриев, О.И. Пантюхин, Г.А. Рябов, Б.В. Солодухин. — EDN WVDMKC // Актуальные проблемы инфотелекоммуникаций в науке и образовании : материалы XIII Междунар. науч.-науч.-метод. конф. : в 4 т., Санкт-Петербург, 27–28 февр. 2024 г. — Санкт-Петербург, 2024. — Т. 1. — С. 277–281.
4. Тураев С.Э. Разработка системы обнаружения вредоносного трафика для повышения количества обнаруженных аномалий / С.Э. Тураев, Д.А. Заколдаев. — EDN GSYVEE // Инженерный вестник Дона. — 2024. — № 11(119). — С. 360–373.
5. Бабичева М.В. Применение методов машинного обучения для автоматизированного обнаружения сетевых вторжений / М.В. Бабичева, И.А. Третьяков. — DOI 10.21822/2073-6185-2023-50-1-53-61. — EDN MGBAGF // Вестник Дагестанского государственного технического университета. Технические науки. — 2023. — Т. 50, № 1. — С. 53–61.
6. Макаров Д.А. Обнаружение аномалий в сетевом трафике с помощью метода «Случайный лес» / Д.А. Макаров, А.А. Байкалов. — EDN IDHIAA // Научный аспект. — 2023. — Т. 15, № 6. — С. 1987–1991.
7. Груздев А.В. Прогнозное моделирование в IBM SPSS, R и Python: метод деревьев решений и случайный лес / А.В. Груздев. — Москва : ДМК Пресс, 2018. — 642 с.
8. Рашка С. Машинное обучение с PyTorch и Scikit-Learn / С. Рашка, Ю. Лю, В. Мирджалили. — Астана : Фолиант, 2024. — 688 с.

9. Васильев Юлий. Python для data science / Юлий Васильев. — Санкт-Петербург : Питер, 2023. — 272 с.
10. Серрано Луис. Грокаем машинное обучение / Луис Серрано. — Санкт-Петербург : Питер, 2024. — 512 с.

References

1. Ivanov S.O. A Technique for Creating and Training an Artificial Neural Network to Detect Network Traffic Anomalies. *Informatsionnye tekhnologii = Information Technologies*, 2024, vol. 30, no. 1, pp. 32–41. (In Russian). EDN: NRGZNE. DOI: 10.17587/it.30.32-41.
2. Pozdnyak I.S., Makarov I.S. Attack Detection Models Using Machine Learning Methods. *Vestnik Rossiiskogo novogo universiteta. Seriya: Slozhnye sistemy: modeli, analiz i upravlenie = Vestnik of Russian New University. Series: Complex Systems: Models, Analysis, Management*, 2024, no. 1, pp. 99–109. (In Russian). EDN: MNMSYZ. DOI: 10.18137/RNU.V9187.24.01.P.99.
3. Dmitriev E.A., Pantyukhin O.I., Ryabov G.A., Solodukhin B.V. Analysis and Selection of Significant Characteristics of Network Traffic For Use in Machine Learning. *Materials of International Scientific Conference, Saint Petersburg, February 27–28, 2024*. Saint Petersburg, 2024, vol. 1, pp. 277–281. (In Russian). EDN: WVDMKC.
4. Turaev S.E., Zakoldaev D.A. Development of a Malmical Traffic Detection System to Increase the Number of Detected Anomalies. *Inzhenernyj vestnik Dona = Engineering journal of Don*, 2024, no. 11, pp. 360–373. (In Russian). EDN: GSYVEE.
5. Babicheva M.V., Tretyakov I.A. Application of Machine Learning Methods for Automated Detection of Network Intrusions. *Yuridicheskii vestnik Dagestanskogo gosudarstvennogo universiteta = Law Herald of Dagestan State University*, 2023, vol. 50, no. 1, pp. 53–61. (In Russian). EDN: MGBAGF. DOI: 10.21822/2073-6185-2023-50-1-53-61.
6. Makarov D.A. Detecting Anomalies in Network Traffic Using Random Forest. *Nauchnyi aspekt = Scientific aspect*, 2023, vol. 15, no. 6, pp. 1987–1991. (In Russian). EDN: IDHIAA.
7. Gruzdev A.V. *Predictive Modeling in IBM SPSS, R and Python: Decision Trees and Random Forest*. Moscow, DMK Press Publ., 2018. 642 p.
8. Raschka S., Lyu Yu., Mirjalili V. *Machine Learning with PyTorch and Scikit-Learn. Develop machine learning and deep learning models with Python*, 2022. 770 p. (Russ. ed.: Rashka S., Lyu Yu., Mirdzhalili V. *Machine Learning with PyTorch and Scikit-Learn*. Astana, Foliant Publ., 2024. 688 p.).
9. Vasil'ev Yulii. *Python for Data Science*. Saint Petersburg, Piter Publ., 2023. 272 p.
10. Serrano L. *Grokking Machine Learning*, 2021. 511 p. (Russ. ed.: Serrano L. *Grokking Machine Learning*. Saint Petersburg, Piter Publ., 2024. 512 p.).

Информация об авторах

Баженов Михаил Дмитриевич — магистрант, кафедры математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: bazhenovmd@yandex.ru.

Бусько Михаил Михайлович — кандидат технических наук, доцент, кафедры математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: BuskoMM@bgu.ru.

Information about the Authors

Mikhail D. Bazhenov — Master's Degree Student, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: bazhenovmd@yandex.ru.

Mikhail M. Busko — PhD in Technical Sciences, Associate Professor, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: BuskoMM@bgu.ru.

Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the Authors

The authors contributed equally to this article. The authors declare no conflicts of interests.

Для цитирования

Баженов М.Д. Исследование достоверности обучающих наборов данных анализатора сетевого трафика / М.Д. Баженов, М.М. Бусько. — DOI 10.17150/2713-1734.2025.7(2).153-164. — EDN FIEARZ // *System Analysis & Mathematical Modeling*. — 2025. — Т. 7, № 2. — С. 153–164.

For Citation

Bazhenov M.D., Busko M.M. Research of the Reliability of Network Traffic Analyzer Training Datasets. *System Analysis & Mathematical Modeling*, 2025, vol. 7, no. 2, pp. 153–164. (In Russian). EDN: FIEARZ. DOI: 10.17150/2713-1734.2025.7(2).153-164.