Современная наука и инновации. 2025. № 2. С. 19-31
ТЕХНИЧЕСКИЕ НАУКИ
ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

Modern Science and Innovations. 2025;(2):19-31
TECHNICAL SCIENCE
INFORMATICS, COMPUTER ENGINEERING AND MANAGEMENT

Научная статья
УДК 615.07:004.8
https://doi.org/10.37493/2307-910X.2025.2.2

Исследование механизмов детектирования кибератак с помощью технологий машинного обучения

Юлия Алексеевна Андрусенко[1], Дмитрий Сергеевич Шавло[2*], Алексей Павлович Плетухин[3], Елена Романовна Семиколеннова[4], Михаил Ильич Кондрашов[5]

[1, 2, 3, 4] Северо-Кавказский федеральный университет (д. 1, ул. Пушкина, Ставрополь, 355017, Россия)
[5] МИРЭА – Российский технологический университет (д. 78, пр. Вернадского, Москва, 119454, Россия)
[1] iuandrusenko@ncfu.ru
[2] ds357@ro.ru
[3] mrster03@gmail.com
[4] elena291204@mail.ru
[5] kondr-mih@mail.ru
*Автор, ответственный за переписку

Аннотация. *В условиях растущей сложности киберугроз и увеличения частоты кибератак актуальность разработки эффективных инструментов для автоматического обнаружения вторжений в информационные системы неуклонно возрастает. Традиционные методы защиты, основанные на сигнатурном анализе, часто не справляются с новыми, неизвестными атаками, что стимулирует исследование альтернативных подходов, в том числе методов машинного обучения (ML). В статье рассматривается применение ML для решения задачи бинарной классификации сетевой активности на нормальную и аномальную (кибератаку) с использованием программной платформы KNIME, ориентированной на визуальное проектирование аналитических пайплайнов.Исследование проводилось на общедоступном датасете Cybersecurity Intrusion Detection, включающем 11 признаков, характеризующих сетевой трафик и пользовательское поведение (время активности, протокол передачи данных, размер пакетов, количество подключений, частота запросов и т.д.). Целью работы стало сравнительное изучение эффективности пяти алгоритмов ML: Decision Trees (деревья решений), Naive Bayes (наивный байесовский классификатор), Random Forest (случайный лес), Gradient Boosted Trees (градиентный бустинг на деревьях) и Simple Regression Tree (простое регрессионное дерево). Выбор моделей обусловлен их распространенностью в задачах обнаружения аномалий и различиями в принципах работы: от простых решающих правил (Decision Trees) до ансамблевых методов (Random Forest, Gradient Boosted Trees), объединяющих несколько "слабых" моделей для повышения точности.Для подготовки данных к обучению применялся метод главных компонент (PCA), позволивший сократить размерность признакового пространства с 11 до 3 компонент без существенной потери информации. Это важный шаг, так как избыточность или коррелированность признаков может негативно влиять на качество моделей. Настройка гиперпараметров (например, глубины деревьев, количества деревьев в Random Forest, скорости обучения в Gradient Boosted Trees) и борьба с переобучением осуществлялись через кросс-валидацию, что обеспечило стабильность результатов на новых данных. Эксперименты показали, что наибольшую точность (83.055%) и площадь под ROC-кривой (AUC = 0.811) продемонстрировали алгоритмы Random Forest и Gradient Boosted Trees. Эти результаты объясняются способностью ансамблевых методов учитывать нелинейные зависимости в данных и устойчивостью к шуму, что критично для задач кибербезопасности, где атаки могут*

маскироваться под нормальную активность. Модели Decision Trees и Simple Regression Tree показали более низкие метрики (точность ~75-78%), что связано с их склонностью к переобучению на небольших датасетах. Naive Bayes, предполагающий независимость признаков, также уступил ансамблевым методам, что подтверждает ограниченность предположения о независимости в контексте сетевых данных.Особое внимание уделено этапам работы: от загрузки и визуализации данных (анализ распределения классов, корреляций между признаками) до обучения моделей и интерпретации результатов. Использование KNIME упростило реализацию пайплайна: платформа предоставляет визуальные инструменты для предобработки данных, настройки моделей и оценки их качества, что делает подход доступным для специалистов в области кибербезопасности без глубоких знаний программирования.Результаты исследования вносят вклад в развитие практических приложений ML для защиты информационных систем. Они демонстрируют, что ансамблевые методы, такие как Random Forest и Gradient Boosted Trees, могут эффективно применяться для обнаружения кибератак в режиме реального времени, особенно в условиях ограниченного набора данных. Перспективными направлениями дальнейших работ являются расширение датасета за счет включения новых типов атак (например, Advanced Persistent Threats, IoT-атак, атак на криптографические протоколы), а также интеграция методов глубокого обучения (например, рекуррентных нейронных сетей для анализа последовательностей сетевых событий) для повышения точности и адаптивности систем обнаружения вторжений.

Ключевые слова: KNIME, машинное обучение, датасет, кибератака, сетевая активность

Для цитирования: *Андрусекно Ю. А., Шавло Д. С., Плетухин А. П., Семиколеннова Е. Р., Кондрашев М. И. Исследование механизмов детектирования кибератак с помощью технологий машинного обучения // Современная наука и инновации. 2025. № 2. С. 19–31. https://doi.org/10.37493/2307-910X.2025.2.2*

Research article

Investigation of cyberattack detection mechanisms using machine learning technologies

Yulia A. Andrusenko[1], Dmitry S. Shavlo[2*], Alexey P. Pletukhin[3], Elena R. Semikolennova[4], Mikhail I. Kondrashov[5]

[1, 2, 3, 4] North-Caucasus Federal University (1, Pushkin St., Stavropol, 355017, Russia)
[5] MIREA - Russian Technological University (78, Vernadsky Ave., Moscow, 119454, Russia)
[1] iuandrusenko@ncfu.ru
[2] ds357@ro.ru
[3] mrster03@gmail.com
[4] elena291204@mail.ru
[5] kondr-mih@mail.ru
*Corresponding author

Abstract. *In the context of the growing complexity of cyber threats and the increasing frequency of cyberattacks, the relevance of developing effective tools for automatic detection of intrusions into information systems is steadily increasing. Traditional protection methods based on signature analysis often fail to cope with new, unknown attacks, which stimulates the study of alternative approaches, including machine learning (ML) methods. The article discusses the use of ML to solve the problem of binary classification of network activity into normal and abnormal (cyberattack) using the KNIME software platform, which is focused on the visual design of analytical pipelines. The study was conducted on the publicly available Cybersecurity Intrusion Detection dataset, which includes 11 features characterizing network traffic and user behavior (activity time, data transfer protocol, packet size, number of connections, request frequency, etc.). The aim of the work was a comparative study of the efficiency of five ML algorithms: Decision Trees, Naive Bayes, Random Forest, Gradient Boosted Trees, and Simple Regression Tree. The choice of models is due to their prevalence in anomaly detection problems and differences in operating principles: from simple decision rules (Decision Trees) to ensemble methods (Random Forest, Gradient Boosted Trees) that combine several "weak" models to*

improve accuracy. To prepare the data for training, the principal component analysis (PCA) was used, which allowed us to reduce the dimensionality of the feature space from 11 to 3 components without significant loss of information. This is an important step, since redundancy or correlation of features can negatively affect the quality of the models. Tuning hyperparameters (e.g. tree depth, number of trees in Random Forest, learning rate in Gradient Boosted Trees) and combating overfitting were performed through cross-validation, which ensured the stability of the results on new data. Experiments showed that the highest accuracy (83.055%) and area under the ROC curve (AUC = 0.811) were demonstrated by the Random Forest and Gradient Boosted Trees algorithms. These results are explained by the ability of ensemble methods to take into account nonlinear dependencies in data and their resistance to noise, which is critical for cybersecurity tasks, where attacks can be disguised as normal activity. Decision Trees and Simple Regression Tree models showed lower metrics (accuracy ~75-78%), which is due to their tendency to overfitting on small datasets. Naive Bayes, which assumes independence of features, also gave way to ensemble methods, which confirms the limitations of the independence assumption in the context of network data. Particular attention is paid to the stages of work: from loading and visualizing data (analysis of class distribution, correlations between features) to training models and interpreting the results. The use of KNIME simplified the implementation of the pipeline: the platform provides visual tools for data preprocessing, model tuning and assessing their quality, which makes the approach accessible to cybersecurity specialists without deep programming knowledge. The results of the study contribute to the development of practical applications of ML for protecting information systems. They demonstrate that ensemble methods such as Random Forest and Gradient Boosted Trees can be effectively used to detect cyberattacks in real time, especially in conditions of a limited data set. Promising areas for further work include expanding the dataset to include new types of attacks (e.g. Advanced Persistent Threats, IoT attacks, attacks on cryptographic protocols), as well as integrating deep learning methods (e.g. recurrent neural networks for analyzing network event sequences) to improve the accuracy and adaptability of intrusion detection systems.

Keywords: KNIME, Machine Learning, Dataset, Cyber Attack, Network Activity

***For citation**: Andrusenko YA, Shavlo DS, Pletukhin AP, Semikolenova ER, Kondrashev MI. Investigation of cyberattack detection mechanisms using machine learning technologies. odern Science and Innovations. 2025;(2):19-31. (In Russ.). https://doi.org/10.37493/2307-910X.2025.2.2*

Введение. В современном цифровом мире кибербезопасноть [1] становится одной из ключевых задач для организаций и предприятий любого масштаба. С ростом числа интернет-пользователей и развитием технологий существенно увеличивается количество кибератак [2], которые становятся более изощрёнными и сложными для обнаружения. По данным различных исследований, ежегодный ущерб от киберпреступлений [3] исчисляется миллиардами долларов, а количество атак продолжается расти.

Традиционные методы защиты, такие как антивирусы и межсетевые экраны, часто не справляются с быстро меняющимися угрозами. Это делает актуальным применение современных технологий, в частности методов машинного обучения, для анализа сетевого трафика и выявления потенциальных атак. Машинное обучение позволяет автоматически обнаруживать аномалии и паттерны, характерные для различных типов кибератак, что значительно повышает эффективность систем обнаружения вторжений.

В данной работе рассматривается применение платформы KNIME для создания моделей машинного обучения, направленных на обнаружение кибервторжений. Исследование основывается на анализе датасета Cybersecurity Intrusion Detection, содержащего информацию о сетевой активности и различных параметрах, которые могут указывать на подозрительное или вредоносное поведение.

Актуальность исследования подтверждается многочисленными научными работами в данной области. Например, в статье [4] авторы исследуют различные подходы машинного обучения для идентификации характерных паттернов brute force атак в сетевом трафике. В работе [5] рассказывается про разработку системы PassREfinder, которая предсказывает риск атак credential stuffing (массовый перебор аккаунтов) путем анализа повторного использования паролей между веб-сайтами с помощью графового представления. В исследовании [6] представлено применении методов машинного

обучения для обнаружения DDoS-атак на устройствах Интернета вещей (IoT) потребительского уровня.

Целью работы является сравнительный анализ различных алгоритмов машинного обучения и определение оптимальных параметров их работы для достижения максимальной точности обнаружения атак. Особое внимание уделяется предобработке данных, настройке гиперпараметров моделей и методам борьбы с переобучением.

Результаты данного исследования могут быть полезны для специалистов в области кибербезопасности при разработке систем обнаружения вторжений, а также для дальнейших исследований в области применения машинного обучения в кибербезопасности. В работе рассматривается различные модели машинного обучения, включая Random Forest, Decision Tree, Gradient Boosted Trees, Simple Regression Tree и Naive Bayes, с последующим детальным анализом их эффективности в обнаружении атак и оптимальной настройке параметров работы.

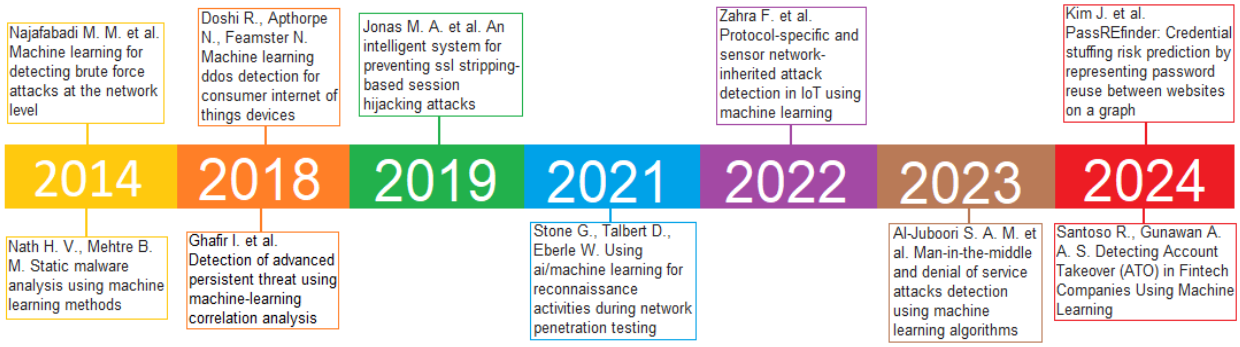

Рисунок 1. Таймлайн исследований в области кибербезопасности /
Figure 1. Cybersecurity Research Timeline

В статье [4] авторы исследуют различные подходы машинного обучения для идентификации характерных паттернов brute force атак в сетевом трафике. В исследовании [5] рассказывается про разработку системы PassREfinder, которая предсказывает риск атак credential stuffing (массовый перебор аккаунтов) путем анализа повторного использования паролей между веб-сайтами с помощью графового представления. В работе [6] представлено применении методов машинного обучения для обнаружения DDoS-атак на устройствах Интернета вещей (IoT) потребительского уровня. В статье [7] говорится о разработке интеллектуальной системы для предотвращения атак, основанных на перехвате сессий путем удаления SSL-шифрования (SSL stripping). В работе [8] автор исследует применения алгоритмов машинного обучения для обнаружения атак типа "человек посередине" (Man-in-the-Middle) и отказа в обслуживании (Denial of Service). В исследовании [9] рассматривается применение искусственного интеллекта и машинного обучения для проведения разведывательных мероприятий во время тестирования на проникновение в компьютерные сети. В статье [10] исследуются методы машинного обучения для выявления атак на захват учетных записей (Account Takeover, ATO) в финтех-компаниях. В работе [11] автор анализирует использование методов машинного обучения и корреляционного анализа для обнаружения сложных угроз типа APT (Advanced Persistent Threat). В статье [12] описывается применение методов машинного обучения для статического анализа вредоносного программного обеспечения (Malware), с целью повышения эффективности его обнаружения и классификации. В исследовании [13] рассматривается использование методов машинного обучения для обнаружения атак, специфичных для протоколов и отличающихся от традиционных сетевых атак, в IoT-системах.

Разбор датасета Cybersecurity Intrusion Detection. Cybersecurity Intrusion Detection Dataset[14] предназначен для обнаружения кибер-вторжений на основе сетевого трафика и поведения пользователей. Каждая запись представляет собой сетевую активность с характеристиками, которые могут указывать на подозрительное или вредоносное поведение.

Исходный датасет содержит 11 столбцов, и в каждом из которых содержится определённые типы данных. В них так же содержится определённая информация, которая может помочь во время обучения моделей.

В ходе исследования проведён анализ данных датасета Cybersecurity Intrusion Detection, представленный в Таблице 1.

Таблица 1. Анализ данных в датасете сетевого трафика /
Table 1. Data analysis in the network traffic dataset

Характеристика	Тип данных	Описание	Диапазон	Примечание
Session id	категориальный	Уникальный идентификатор сессии	Уникальный идентификатор сессии	Используется для однозначной идентификации записей
Network packet size	количественный	Размер сетевого пакета	Размер сетевого пакета	Может указывать на подозрительную активность (например, аномально большие или малые пакеты)
Protocol type	категориальный	Тип протокола	TCP (Transmission Control Protocol), UDP (User Datagram Protocol) и ICMP (Internet Control Message Protocol)	Тип сетевого протокола может подсказать на конкретную атаку, так как некоторые протоколы чаще используются в атаках
Login attempts	количественный	Количество попыток входа в систему	От 1 до 12	Высокое значение может указывать на брутфорс-атаку
Session duration	количественный	Продолжительность сессии (в секундах)	От 1.37 до 5443.23	Долгие сессии могут быть признаком сканирования сети или скрытой активности
Encryption used	категориальный	Тип шифрования	DES, AES, None	Отсутствие шифрования (None) повышает писк перехвата данных
Ip reputation score	количественный	Репутация IP-адреса	оценка от 0 до 1	Чем ближе к 0, тем выше вероятность, что IP связан с вредоносной активностью
Failed logins	количественный	Количество неудачных попыток входа	От 0 до 5	Показатель подозрительной активности (например, подбор паролей)
Browser type	категориальный	Используемый браузер	Edge, Firefox, Chrome, Safari, Unknown	Некоторые браузеры могут быть связаны с уязвимостями или автоматизированными ботами
Unusual time access	бинарный	Метка доступа в необычное время	0 — нет, 1 — да	Помогает выявить аномальную активность (например, ночные запросы)
Attack detected	бинарный	Целевая переменная: обнаружена ли атака	0 — нет, 1 — да	–

Датасет предназначен для задач классификации в области кибербезопасности, где модель машинного обучения обучается предсказывать наличие атаки на основе остальных параметров. Признаки включают как количественные, так и категориальные данные, что требует предобработки перед использованием в алгоритмах.

Исследование разных моделей машинного обучения для работы с датасетом. Для исследования моделей будем использовать платформу KNIME (Konstanz Information Miner). Платформа используется для анализа и визуализации данных и построения моделей машинного обучения

KNIME – это открытая платформа для работы с анализом данных, машинным обучением и созданием аналитических процессов. Программа имеет графический интерфейс, который позволяет пользоваться машинным обучением, не имея знаний программирования, используя ноды (блоки) вместо кода. Работа с данной платформой станет основной задачей данной научно-исследовательской работы.

Все модели, будут строиться по одному алгоритму. Обобщающая схема модели обучения с обучающими блоками приведена на рис. 2. Блок ML (Машинное обучение) в каждой модели заменяется на разные блоки.

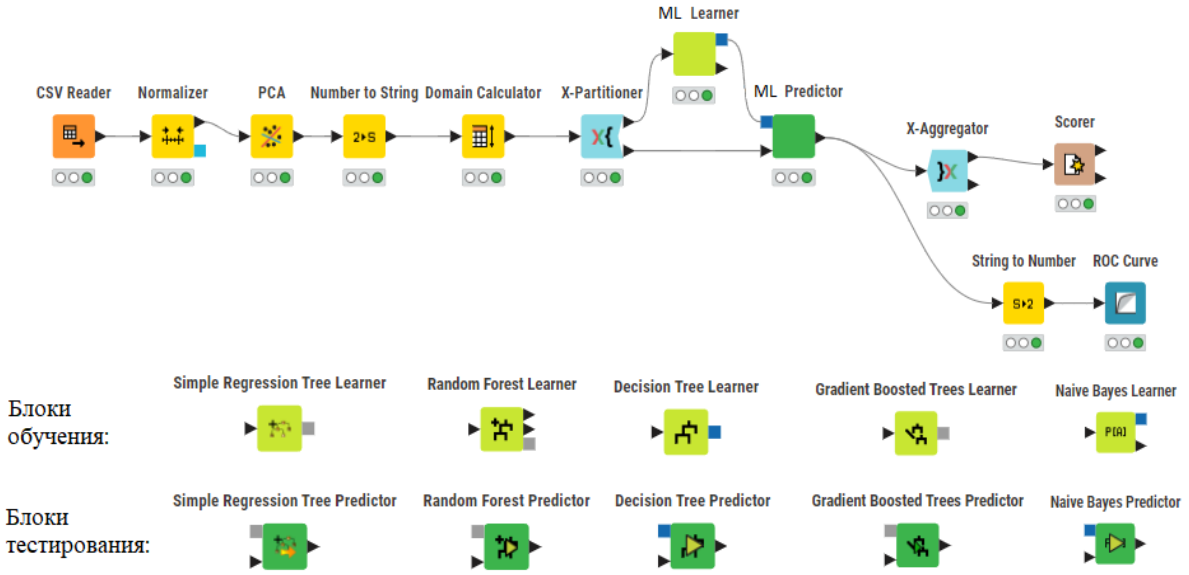

Рисунок 2. Обобщающая схема модели машинного обучения /
Figure 2. Generalized diagram of a machine learning model

В данной обобщающей схеме (таблица 2) задействовано 12 узлов, которые помогут обучить машину и выяснить точность её прогнозирования:

Таблица 2. Предназначения узлов в каждой модели /
Table 2. Purpose of the nodes in each model

CSV Reader	**загружает CSV-файл**
Normalizer	нормализует данные
PCA (Principal Component Analysis)	уменьшает размерность данных
Number to String	переводит числовые значения в строковые
Domain Calculator	сканирование данных и обновление списка возможных значений и/или минимальных и максимальных значений выбранных столбцов

X-Partitioner	делит данные на обучающие и тестовые и осуществляет кросс-валидацию
ML Learner	обучает модель
ML Predictor	тестирует модель
X-Aggregator	осуществляет кросс-валидацию
Scorer	выводит точность модели
String to Number	переводит строковые значения в числа
ROC Curve (Receiver Operating Characteristic Curve)	выводит ROC-кривую

PCA (Метод главных компонент) – это метод, при котором сокращается размерность данных, сохраняя при этом наибольшее количество информации [15].

ROC-кривая (Рабочая характеристика приёмника) — график, который позволяет оценить качество бинарной классификации [16].

Исследовано 5 моделей, каждая из которых использовала свой метод машинного обучения. Среди них: Дерево решений [17], Случайный лес [18], Простое дерево регрессии [19], Градиентный бустинг деревьев[20] и Наивный Байес [21].

Каждая модель показала следующие результаты точности при настройках по умолчанию (табл. 3).

Таблица 3. Точность моделей машинного обучения на изначальных настройках / Table 3. Accuracy of machine learning models at initial settings

Метод машинного обучения	**Точность (в %)**
Случайный лес	82,93
Дерево решений	81,113
Градиентный бустинг деревьев	83,045
Простое дерево регрессии	78,819
Наивный Байес	81,975

Чтобы повысить точность моделей, необходимо провести дополнительные настройки в них. Одна из таковых является блок PCA (табл. 4).

Таблица 4. Зависимость точности моделей от значения блока PCA / Table 4. Dependence of model accuracy on the value of the PCA block

Значение блока PCA	**Точность модели (в %)**				
	Случайный лес	**Дерево решений**	**Градиентный бустинг деревьев**	**Простое дерево регрессии**	**Наивный Байес**
1	82,94	81,609	83,014	79,04	82,007
2	82,961	81,609	83,014	79,029	81,996
3	83,003	81,76	83,003	78,809	81,996
4	82,898	81,874	83,055	79,04	82,017
5	82,951	81,775	83,076	79,239	82,038
6	83,014	81,478	83,003	78,368	81,986

Точность моделей от значения PCA более наглядно представлена на рисунке 3.

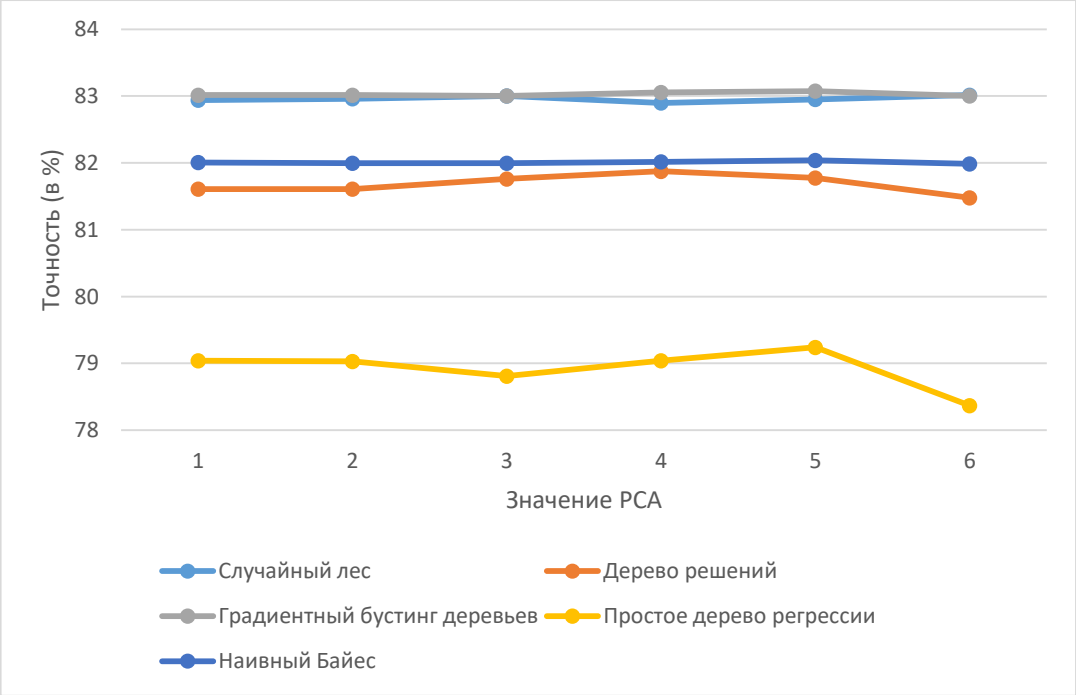

Рисунок 3. Зависимость точности моделей от значения блока PCA /
Figure 3. Dependence of model accuracy on the value of the PCA block

Из данных результатов можно сделать вывод о том, что для лучшей точности нужно сократить колонки: Случайный лес – до 6 колонок, Дерево решений – до 4 колонок, Градиентный бустинг деревьев – до 5 колонок, Простое дерево регрессии – до 5 колонок, Наивный Байес – до 5 колонок.

Для каждой модели была проведена настройка глубины обучения с теми количествами колонок, которые были получены до этого (см. Табл. 5).

Таблица 5. Зависимость точности моделей от глубины обучения /
Table 5. Dependence of model accuracy on training depth

Глубина обучения	Точность модели (в %)				
	Случайный лес	Дерево решений	Градиентный бустинг деревьев	Простое дерево регрессии	Наивный Байес
1	58,006	81,1	81,944	33,606	55,29
2	76,146	81,332	83,055	51,997	55,29
3	82,982	81,092	83,055	74,74	73,44
4	83,024	81,444	83,034	79,291	73,44
5	83,055	81,902	82,898	79,501	73,44
6	83,055	81,003	82,72	79,071	73,44
7	83,034	81,092	82,374	79,511	82,017
8	83,024	81,003	81,86	79,27	82,028
9	83,014	81,241	81,65	79,281	81,986
10	82,993	81,092	81,546	79,27	82,007

Из данных результатов можно сделать вывод о том, что модели обучения достигают наивысшего порядка при определённой глубине обучения: Случайный лес – 5-6 глубина, Дерево решений – 5 глубина, Градиентный бустинг деревьев – 2-3 глубина, Простое дерево регрессии – 7 глубина, Наивный Байес – 8 глубина (рис. 4).

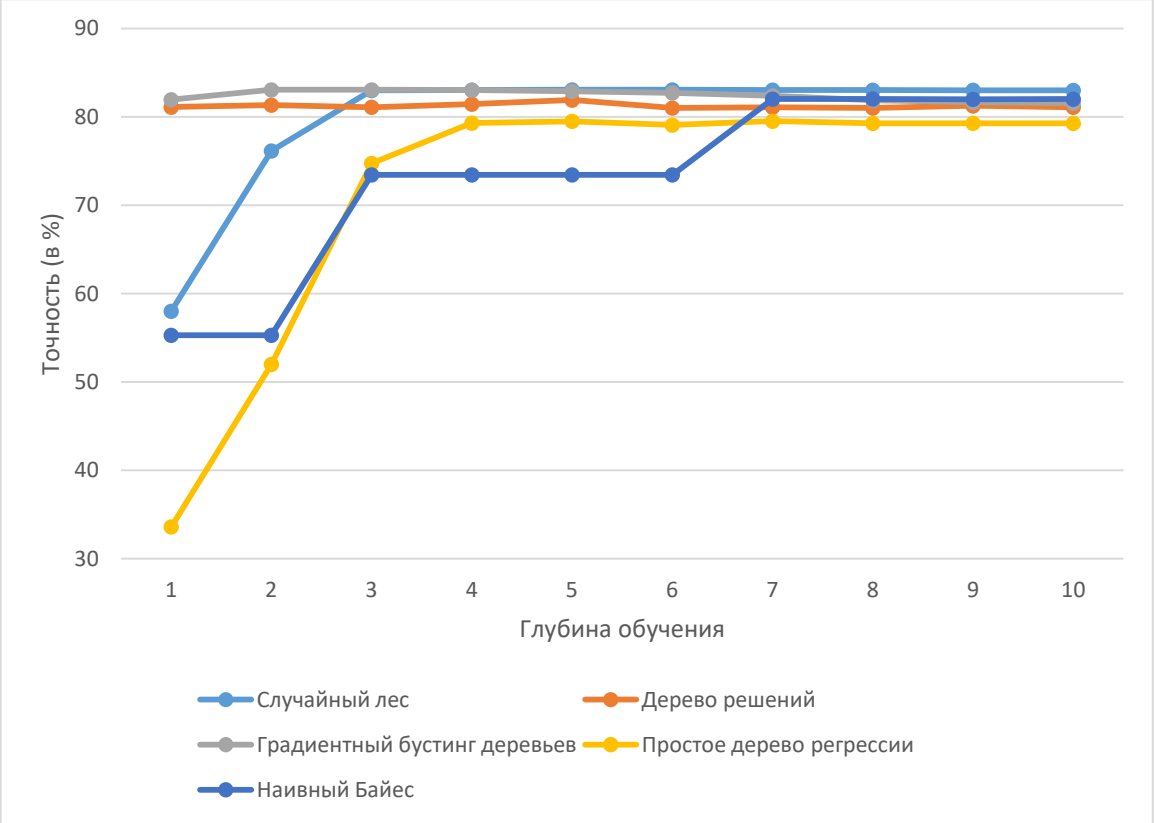

Рисунок 4. Зависимость точности моделей от глубины обучения (в %) /
Figure 4. Dependence of model accuracy on training depth (in %).

При машинном обучении может случится переобучение модели. Это происходит, когда модель начинает запоминать обучающие данные, что приводит к высокой точности, но если будут ранее неизвестные данные, то точность будет крайне низкая. Один из методов решения этой проблемы является кросс-валидация – метод оценки производительности модели обучения, используя для проверки способность обобщать новые уникальные данные [22].

Таблица 6. Лучшие результаты моделей /
Table 6. The best results of the models

Метод машинного обучения	**Значение PCA**	**Глубина обучения**	**Наивысшая точность, %**
Случайный лес	6	5-6	83,055
Дерево решений	4	5	81,902
Градиентный бустинг деревьев	5	2-3	83,055
Простое дерево регрессии	5	7	79,511
Наивный Байес	5	8	82,028

Для того чтобы увидеть качество прогнозирования каждой модели, нужно воспользоваться ROC-кривыми. Они показывают зависимость между чувствительностью (True Positive Rate, TPR) и специфичностью (False Positive Rate, FPR) модели при изменении порога принятия решений.

AUC (площадь под кривой) используется как числовая метрика качества модели:• AUC = 1 – идеальная модель.• AUC = 0.5 – модель начинает случайно гадать, что делает её ненадёжной• AUC < 0.5 – модель работает плохо (её предсказания противоположны истинным)Исходя из графиков ROC-кривых (рис. 5) Random Forest и Gradient Boosted Trees имеют одинаковый максимальный AUC = 0.811 и демонстрируют наилучшую способность различать классы.

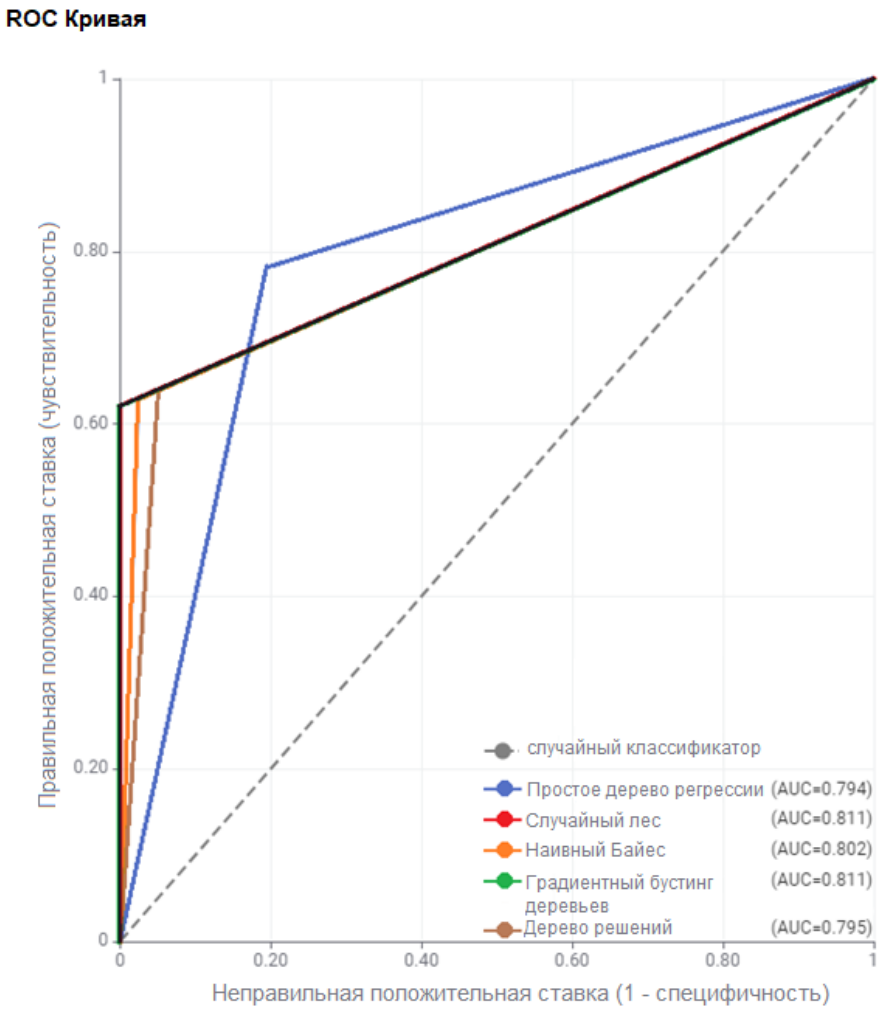

Рисунок 5. Графики ROC-кривых /
Figure 5. ROC curve graphs

Они являются наиболее предпочтительными для выявления кибератак на основе анализа сетевого трафика и поведения пользователей.

Заключение. В ходе исследования были рассмотрены различные методы машинного обучения для обнаружения кибератак с использованием платформы KNIME. Был проведён сравнительный анализ эффективности пяти алгоритмов на основе датасета Cybersecurity Intrusion Detection.

Анализ результатов показал, что наибольшую точность демонстрируют модели Случайный лес и Градиентный бустинг деревьев, достигая значения AUC=0,811 при оптимальных настройках. Для достижения максимальной производительности моделей потребовалось выполнить тщательную настройку параметров: количество главных компонентов в PCA и глубину обучения деревьев.

Исследование подтвердило перспективность применения машинного обучения в системах обнаружения вторжений. Разработанная методика в KNIME может быть использована для создания эффективных систем защиты от киберугроз путём анализа сетевого трафика и поведения пользователей.

Полученные результаты представляют практическую ценность для специалистов по кибербезопасноти и могут служить основой для дальнейшего совершенствования систем обнаружения атак.

ЛИТЕРАТУРА

1. Kemmerer R. A. Cybersecurity // 25th International Conference on Software Engineering, 2003. Proceedings. IEEE, 2003. P. 705–715.
2. Biju J. M., Gopal N., Prakash A. J. Cyberattacks and its different types // International Research Journal of Engineering and Technology. 2019. Vol. 6. No. 3. P. 4849–4852.
3. Sabillon R., Cano M. J. J., Serra-Ruiz J., Cavaller V. Cybercrime and cybercriminals: A comprehensive study // International Journal of Computer Networks and Communications Security. 2016. Vol. 4. No. 6. P. 165–176.
4. Najafabadi M. M., Khoshgoftaar T. M., Kemp C., Seliya N., Zuech R. Machine learning for detecting brute force attacks at the network level // 2014 IEEE International Conference on Bioinformatics and Bioengineering. Boca Raton: IEEE, 2014. P. 379–385.
5. Kim J., Song M., Seo M., Jin Y., Shin S. PassREfinder: Credential stuffing risk prediction by representing password reuse between websites on a graph // 2024 IEEE Symposium on Security and Privacy (SP). San Francisco: IEEE, 2024. P. 1385–1404.
6. Doshi R., Apthorpe N., Feamster N. Machine learning ddos detection for consumer internet of things devices // 2018 IEEE Security and Privacy Workshops (SPW). IEEE, 2018. P. 29–35.
7. Jonas M. A., Hossain M. S., Islam R., Narman H. S., Atiquzzaman M. An intelligent system for preventing ssl stripping-based session hijacking attacks // MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM). Norfolk: IEEE, 2019. P. 1–6.
8. Al-Juboori S. A. M., Hazzaa F., Jabbar Z. S., Salih S., Gheni H. M. Man-in-the-middle and denial of service attacks detection using machine learning algorithms // Bulletin of Electrical Engineering and Informatics. 2023. Vol. 12. No. 1. P. 418–426.
9. Stone G., Talbert D., Eberle W. Using Using AI/Machine Learning for Reconnaissance Activities During Network Penetration Testing // International Conference on Cyber Warfare and Security. Academic Conferences International Limited, 2021. P. 541–XIV.
10. Santoso R., Gunawan A. A. S. Detecting Account Takeover (ATO) in Fintech Companies Using Machine Learning // 2024 6th International Conference on Cybernetics and Intelligent System (ICORIS). Surakarta: IEEE, 2024. P. 1–6.
11. Ghafir I., Hammoudeh M., Prenosil V., Han L., Hegarty R., Rabie K., Aparicio-Navarro F. J. Detection of advanced persistent threat using machine-learning correlation analysis // Future Generation Computer Systems. 2018. Vol. 89. P. 349–359.
12. Nath H. V., Mehtre B. M. Static malware analysis using machine learning methods // International Conference on Security in Computer Networks and Distributed Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. P. 440–450.
13. Zahra F., Jhanjhi N. Z., Khan N. A., Brohi S. N., Masud M., Aljahdali S. Protocol-specific and sensor network-inherited attack detection in IoT using machine learning // Applied Sciences. 2022. Vol. 12. No. 22. P. 11598.
14. Samudrala D. N. K. Cybersecurity Intrusion Detection Dataset // Kaggle. [Electronic resource]. URL: https://www.kaggle.com/datasets/dnkumars/cybersecurity-intrusion-detection-dataset (accessed: 15.03.2025).
15. Mackiewicz A., Ratajczak W. Principal components analysis (PCA) // Computers and Geosciences. 1993. Vol. 19. No. 3. P. 303–342.
16. Nahm F. S. Receiver operating characteristic curve: overview and practical use for clinicians // Korean journal of anesthesiology. 2022. Vol. 75. No. 1. P. 25–36.
17. Charbuty B., Abdulazeez A. Classification based on decision tree algorithm for machine learning // Journal of applied science and technology trends. 2021. Vol. 2. No. 1. P. 20–28.
18. Rigatti S. J. Random forest // Journal of Insurance Medicine. 2017. Vol. 47. No. 1. P. 31–39.
19. Loh W. Y. Classification and regression trees // Wiley interdisciplinary reviews: data mining and knowledge discovery. 2011. Vol. 1. No. 1. P. 14–23.
20. Ye J., Chow J. H., Chen J., Zheng Z. Stochastic gradient boosted distributed decision trees // Proceedings of the 18th ACM conference on Information and knowledge management. 2009. P. 2061–2064.

21. Webb G. I., Keogh E., Miikkulainen R. Naïve Bayes // Encyclopedia of machine learning. 2010. Vol. 15. No. 1. P. 713–714.
22. Berrar D. Cross-validation. Encyclopedia of Bioinformatics and Computational Biology. 2019. Vol. 1. P. 542–545.

REFERENCES

1. Kemmerer RA. Cybersecurity. 25th International Conference on Software Engineering, 2003. Proceedings. IEEE, 2003. P. 705-715.
2. Biju JM., Gopal N, Prakash AJ. Cyberattacks and its different types. International Research Journal of Engineering and Technology. 2019;6(3):4849-4852.
3. Sabillon R, Cano MJJ, Serra-Ruiz J, Cavaller V. Cybercrime and cybercriminals: A comprehensive study. International Journal of Computer Networks and Communications Security. 2016;4(6):165-176.
4. Najafabadi MM., Khoshgoftaar TM, Kemp C, Seliya N, Zuech R. Machine learning for detecting brute force attacks at the network level. 2014 IEEE International Conference on Bioinformatics and Bioengineering. Boca Raton: IEEE, 2014. P. 379-385.
5. Kim J, Song M, Seo M, Jin Y, Shin S. PassREfinder: Credential stuffing risk prediction by representing password reuse between websites on a graph. 2024 IEEE Symposium on Security and Privacy (SP). San Francisco: IEEE, 2024. P. 1385-1404.
6. Doshi R, Apthorpe N, Feamster N. Machine learning ddos detection for consumer internet of things devices. 2018 IEEE Security and Privacy Workshops (SPW). IEEE, 2018. P. 29-35.
7. Jonas MA, Hossain MS, Islam R, Narman HS, Atiquzzaman M. An intelligent system for preventing ssl stripping-based session hijacking attacks. MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM). Norfolk: IEEE, 2019. P. 1-6.
8. Al-Juboori SAM, Hazzaa F, Jabbar ZS, Salih S, Gheni HM. Man-in-the-middle and denial of service attacks detection using machine learning algorithms. Bulletin of Electrical Engineering and Informatics. 2023;12(1):418-426.
9. Stone G., Talbert D., Eberle W. Using AI/Machine Learning for Reconnaissance Activities During Network Penetration Testing. International Conference on Cyber Warfare and Security. Academic Conferences International Limited, 2021. P. 541-XIV.
10. Santoso R, Gunawan AAS. Detecting Account Takeover (ATO) in Fintech Companies Using Machine Learning // 2024 6th International Conference on Cybernetics and Intelligent System (ICORIS). Surakarta: IEEE, 2024. P. 1-6.
11. Ghafir I, Hammoudeh M, Prenosil V, Han L, Hegarty R, Rabie K, Aparicio-Navarro FJ. Detection of advanced persistent threat using machine-learning correlation analysis. Future Generation Computer Systems. 2018;89:349-359.
12. Nath HV, Mehtre BM. Static malware analysis using machine learning methods. International Conference on Security in Computer Networks and Distributed Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. P. 440-450.
13. Zahra F, Jhanjhi NZ, Khan NA, Brohi SN, Masud M, Aljahdali S. Protocol-specific and sensor network-inherited attack detection in IoT using machine learning. Applied Sciences. 2022;12(22):11598.
14. Samudrala DNK. Cybersecurity Intrusion Detection Dataset. Kaggle. Available from: https://www.kaggle.com/datasets/dnkumars/cybersecurity-intrusion-detection-dataset [Accessed 15 March 2025].
15. Mackiewicz A, Ratajczak W. Principal components analysis (PCA). Computers and Geosciences. 1993;19(3):303-342.
16. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean journal of anesthesiology. 2022;75(1):25-36.
17. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of applied science and technology trends. 2021;2(1):20-28.
18. Rigatti SJ. Random forest. Journal of Insurance Medicine. 2017;47(1):31-39.

19. Loh WY. Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery. 2011;1(1):14-23.
20. Ye J, Chow JH, Chen J, Zheng Z. Stochastic gradient boosted distributed decision trees. Proceedings of the 18th ACM conference on Information and knowledge management. 2009. P. 2061-2064.
21. Webb GI, Keogh E, Miikkulainen R. Naive Bayes. Encyclopedia of machine learning. 2010;15(1):713-714.
22. Berrar D. Cross-validation. Encyclopedia of Bioinformatics and Computational Biology. 2019;1:542-545.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Юлия Алексеевна Андрусенко – старший преподаватель, Северо-Кавказский федеральный университет, iuandrusenko@ncfu.ru

Дмитрий Сергеевич Шавло – студент, Северо-Кавказский федеральный университет, ds357@ro.ru

Алексей Павлович Плетухин – студент, Северо-Кавказский федеральный университет, mrster03@gmail.com

Елена Романовна Семиколеннова – студент, Северо-Кавказский федеральный университет, elena291204@mail.ru

Михаил Ильич Кондрашов – студент, МИРЭА – Российский технологический университет, kondr-mih@mail.ru

Вклад авторов: все авторы внесли равный вклад в подготовку публикации.
Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию: 01.04.2025;
одобрена после рецензирования: 13.05.2025;
принята к публикации: 01.06.2025.

INFORMATION ABOUT THE AUTHORS

Yulia A. Andrusenko – Senior Lecturer, North-Caucasus Federal University, iuandrusenko@ncfu.ru

Dmitry S. Shavlo – Student, North-Caucasus Federal University, ds357@ro.ru

Alexey P. Pletukhin – Student, North-Caucasus Federal University, mrster03@gmail.com

Elena R. Semikolennova – Student, North-Caucasus Federal University, elena291204@mail.ru

Mikhail I. Kondrashov – Student, MIREA - Russian Technological University, kondr-mih@mail.ru

Contribution of the authors: the authors contributed equally to this article.
Conflict of interest: the authors declare no conflicts of interests.

The article was submitted: 01.04.2025;
approved after reviewing: 13.05.2025;
accepted for publication: 01.06.2025.