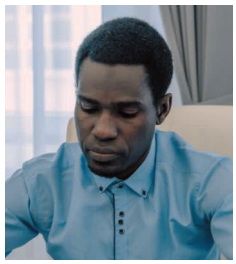
Выбор алгоритма машинного обучения для обнаружения вторжений в IoT

Choosing the machine learning algorithm for detecting intrusions into IoT

Нианг П.М.[1]*, Сидоренко В.Г.[1]
Niang P.M.[1]*, Sidorenko V.G.[1]
[1]РУТ(МИИТ), Москва, Российская Федерация
[1]RUT(MIIT), Moscow, Russian Federation
**malickdiarra30@gmail.com*



Нианг П.М.

Сидоренко В.Г.

Резюме. Цель. *Целью работы является повышение безопасности IoT-устройств путем применения алгоритмов машинного обучения для обнаружения атак в сетях IoT. Актуальность поставленной цели определяется постоянным ростом числа подобных атак в мире и широким распространением систем IoT. В статье приведены соответствующие статистические данные. Анализ имеющихся работ показал, что различные методы рассматривались без связи и сравнения друг с другом, поэтому цель данной работы – определить наиболее перспективный алгоритм машинного обучения для обнаружения атак в сетях IoT – актуальна.* **Методы.** *В статье для обнаружения атак в сетях IoT использовались следующие методы машинного обучения: логистическая регрессия, SVC, «случайный лес», метод K-ближайших соседей, метод k-средних, наивный байесовский классификатор и варианты градиентного бустинга (XGBoost, AdaBoost и CatBoost). Новым является сравнение результатов применения контролируемых алгоритмов с алгоритмом K-means, который является неконтролируемым алгоритмом, для обнаружения атак в сетях IoT. Для обучения создаваемых систем обнаружения атак использовался набор данных UNSWNB15, который содержит данные о девяти видах атак. Количество записей составляет более 80 тысяч. Более половины записей – это записи об атаках. Сравнение методов проводилось по нескольким метрикам.* **Результаты.** *Разработана структура и реализована программно система обнаружения вторжений, включающая этапы от анализа исходных данных до вывода окончательных статистических данных. Результаты показывают, что алгоритм «случайный лес» является лучшим из рассмотренных. Одновременно метод имеет хорошие показатели по быстродействию обучения. Это означает, что данный алгоритм может быть развернут и применен с наибольшим успехом.* **Заключение.** *В этой статье представлены результаты сравнения различных алгоритмов машинного обучения для обнаружения вторжений в устройства IoT. Точность и кривая ROC-AUC используются для оценки эффективности используемых моделей. Сравнивая используемые модели алгоритмов, мы обнаружили, что модель RandomForestClassifier алгоритма Random Forest имеет хорошую точность, самый высокий AUC и быстрое время выполнения, а это означает, что этот алгоритм является наиболее эффективным при обнаружении вторжений в сети IoT. Продолжение исследований связано с различением типа атаки.*

Abstract. Aim. *The paper aims to improve the security of IoT devices by applying machine learning algorithms to detect attacks against IoT networks. The relevance of the goal is defined by the ever-growing number of such attacks around the world and the widespread use of IoT systems. The paper provides relevant statistical data. An analysis of the available papers showed that various methods were examined individually and were not compared to each other, so the aim of this paper that consists in identifying the most promising machine learning algorithm for detecting attacks against IoT networks is of relevance.* **Methods.** *The paper used the following machine learning methods to detect attacks against IoT networks: logistic regression, SVC, random forest, K-nearest neighbour method, k-means method, naive Bayes classifier, and variants of gradient boosting (XGBoost, AdaBoost, and CatBoost). The novelty consists in the comparison of the outputs of the supervised algorithms with the unsupervised K-means in the context of detection of attacks against IoT networks. The attack detection systems under development were trained using the UNSWNB15 dataset that contains data on nine types of attacks. The number of entries is more than 80 thousand. More than half of the entries deal with attacks. The methods were compared using a number of metrics.* **Results.** *An intrusion detection system was structurally defined and implemented. The stages of its operation include the analysis of input data and the output of final statistical data. The results show that the random forest algorithm is the best one out of those examined. The method also performs well in terms of learning speed. That means that the algorithm can be deployed and applied with the greatest success.* **Conclusions.** *This paper presents the*

results of comparing various machine learning algorithms in the context of IoT device intrusion detection. The accuracy and the ROC-AUC curve are used to evaluate the efficiency of the employed models. Having compared the models of the employed algorithms we found that the RandomForestClassifier model has the highest accuracy and a high AUC, which means that this algorithm is the most efficient in terms of IoT network intrusion detection. Further research will be dedicated to distinguishing between the types of attack.

Ключевые слова: *Интернет вещей, обнаружение вторжений, машинное обучение.*
Keywords: *Internet of things, intrusion detection, machine learning.*

Для цитирования: *Нианг П.М., Сидоренко В.Г. Выбор алгоритма машинного обучения для обнаружения вторжений в IoT // Надежность. 2024. №3. С. 44-51. https://doi.org/10.21683/1729-2646-2024-24-3-44-51*
For citation: *Niang P.M., Sidorenko V.G. Choosing the machine learning algorithm for detecting intrusions into IoT. Dependability 2024;3:44-51. https://doi.org/10.21683/1729-2646-2024-24-3-44-51*

Поступила: *29.02.2024* / **После доработки:** *16.05.2024* / **К печати:** *09.09.2024*
Received on: *29.02.2024* / **Revised on:** *16.05.2024* / **For printing:** *09.09.2024*

Введение

Интернет вещей (IoT) соединяет различные устройства, датчики и системы через Интернет, позволяя им обмениваться данными и автоматизировать процессы. За последние годы использование подобных устройств значительно возросло. Этими устройствами можно управлять и получать к ним доступ в любое время и в любом месте через Интернет. IoT можно определить как взаимосвязь между Интернетом и физическими объектами, местами и окружающей средой. Отдельные области IoT касаются транспорта, промышленности, домашней автоматизации, здравоохранения, умного дома и т. д.

В этой статье основное внимание уделяется сетевой безопасности устройств IoT. Эти устройства особенно уязвимы для многих кибератак из-за низкого энергопотребления и низких требований к вычислительным ресурсам, что затрудняет реализацию аутентификации и криптографии в них.

Система обнаружения вторжений (IDS) является решением безопасности для устройств IoT. Обнаружение сетевых вторжений на основе аномалий играет важную роль в защите сетей от различных вредоносных действий. Обнаружение вторжений – это процесс сканирования интеллектуальных устройств и сетевых источников на предмет несанкционированной активности. При взломе сети или устройства безопасность оказывается под угрозой. Из-за незаконных действий злоумышленника система не отвечает на запросы пользователей. IDS предназначена для поддержания безопасности системы путем выявления несанкционированных действий. Обеспечение конфиденциальности, целостности и доступности является важнейшим параметром IDS. Когда происходит вторжение, информация теряет свою конфиденциальность, целостность или доступность.

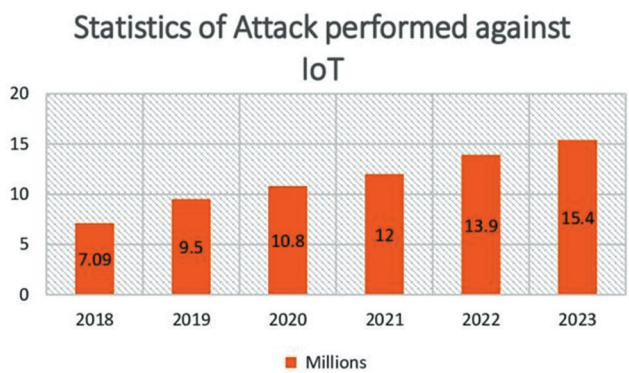


Рис. 1. Статистика атак на IoT [1]

Сетевые атаки на IoT-устройства с каждым годом становятся все более масштабными. На рис. 1 представлена статистика атак на IoT-решения с 2018 по 2023 год [1]. На рис. 2 показаны десять стран и территорий,

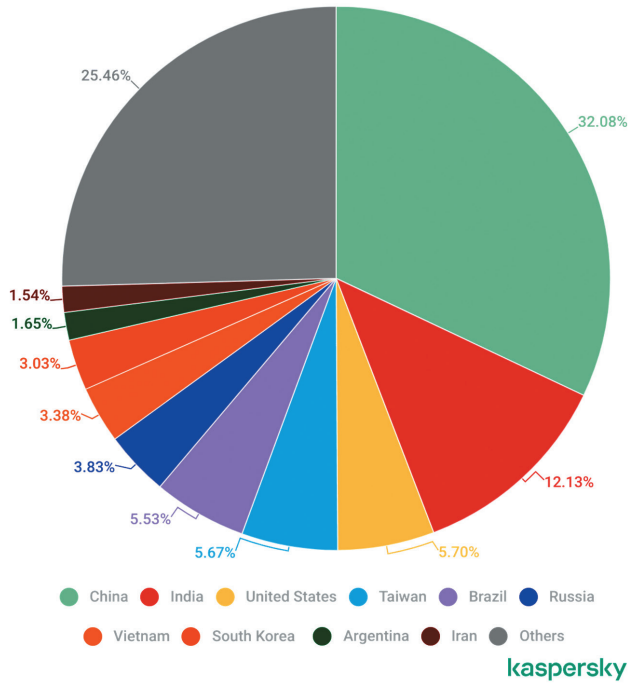


Рис. 2. Десять стран и территорий, где находилось большинство устройств, атаковавших приманки Касперского в 2023 г. [2]

где располагалось большинство атакующих [2]. Данное исследование показывает, что на долю России приходится 3,83%, а Сенегал входит в состав группы стран, обозначенных «Others», на долю которых приходится 25,46% атакующих.

Основная цель IDS – точно идентифицировать атаки с меньшим количеством ложных тревог. Алгоритмы машинного обучения широко используются при разработке IDS благодаря своей скорости, эффективности и высокой производительности [3, 4].

Цель данной работы – определение возможности использования алгоритмов машинного обучения для обнаружения атак в сетях IoT и повышения безопасности IoT-устройств. Статья включает в себя анализ состояния вопроса, синтез архитектуры и методологии предлагаемой системы, тестирование и сравнительный анализ результатов, заключение.

1. Анализ состояния вопроса

Многие исследователи разработали и внедрили различные модели для обнаружения атак в сетях IoT с использованием различных методов.

В работе [4] предложена реализация обнаружения угроз в IoT на основе искусственных нейронных сетей для решения проблем аутентификации. Они использовали контролируемый алгоритм обучения для обнаружения атак, в рамках которого контроллер отклоняет команды после того, как классифицирует их как угрозу. В работе [5] показана возможность применения различных алгоритмов машинного обучения для эффективного обнаружения аномалий в наборе данных о вторжениях в сети IoT. В работе [6] предложена архитектура системы обнаружения сетевых вторжений на основе Ensemble Learning под названием ELNIDS для обнаружения атак маршрутизации на протокол маршрутизации IPv6 для сетей с низким энергопотреблением и потерями. Реализовано четыре различных классификатора машинного обучения на основе ансамблей, включая Boosted Trees, Bagged Trees, Subspace Discriminant и RUSBoosted Trees. Для оценки предложенной модели обнаружения вторжений они использовали набор данных RPL-NIDDS17, который содержит следы пакетов атак. В работе [3] предложена IDS на основе глубокого обучения и машинного обучения для борьбы с атаками безопасности в сетях IoT. В модели обнаружения атак используются сети длинной краткосрочной памяти (LSTM) и метод K-ближайших соседей (KNN), а производительность этих алгоритмов сравнивается друг с другом на основе значений времени обнаружения, статистики Каппа, среднего геометрического и чувствительности. В работе [3] разработана IDS для сред IoT с использованием комбинации методов машинного обучения (наивное дерево Байеса (NBTree), случайное дерево (RandomTree), дерево решений (Decision Tree)), посредством которой пользователь получает предупреждение в случае обнаружения аномалии. В работе [7] выполнено сравнение нескольких алгоритмов глубокого обучения, включая алгоритмы: глубокая нейронная сеть (DNN), сверточная нейронная сеть (CNN), длинная кратковременная память (LSTM) и автоэнкодер (AE), чтобы определить наиболее эффективный алгоритм решения проблем сетевой безопасности с использованием обнаружения вторжений. В исследовании использовался набор данных UNSW-NB15 для тестирования и бинарная классификация для оценки. Результаты показали, что алгоритм DNN достиг значения точности 99,76% и значения потерь 0,006%, превзойдя другие алгоритмы. В работе [8] предложен подход, который включал использование алгоритма «случайный лес» (Random Forest) для уменьшения размерности и выбора оптимального подмножества исходного набора данных. Затем для обнаружения и идентификации вторжений использовался метод ансамблевого обучения. В работе [9] показана важность фильтрации и выборки трафика для оценки надежности устройств Интернета вещей. В работе [10] предложили метод, который реализует IDS, использующую обратное распространение ошибки (BP) вместе с радиальной нейронной сетью с базисной функцией (BFR). В работе [11] представлена работа над IDS путем обнаружения аномалий в сети умного дома с использованием машины экстремального обучения и искусственной иммунной системы (AIS-ELM). AIS использует клональный алгоритм для оптимизации входных параметров, а ELM анализирует входной параметр для лучшей сходимости при обнаружении аномальной активности.

В публикациях Российских авторов тема безопасности IoT также находит отражение и должна учитываться при разработке концепции и критериев оценки технологической независимости и безопасности объектов критической информационной инфраструктуры [12]. Подобные объекты, как правило, являются элементами более крупных систем, объединенных через Интернет. Системы управления городской транспортной системой могут быть отнесены к этому классу систем [13]. Разработан подход к обнаружению аномалий в сетевом трафике путем определения степени самоподобия трафика с помощью фрактального анализа и статистических методов [14]. Полученные результаты по применению методов машинного обучения для прогнозирования отказов объектов транспортных систем также могут служить основой для развития систем безопасности IoT [15, 16].

Таким образом, в существующих работах различные методы рассматривались без связи и сравнения друг с другом. Основная цель данной работы – определить наиболее перспективный алгоритм машинного обучения для обнаружения атак в сетях IoT.

Для этого мы выбираем различные модели алгоритмов (логистическая регрессия, метод опорных векторов (SVC), случайный лес, KNN, метод k-средних (K-means), наивный байесовский классификатор (Naive

Bayes), варианты реализации градиентного бустинга (XGBoost, AdaBoost и CatBoost) и сравниваем их производительность.

Новыми особенностями нашей работы являются:

– достижение точности, превышающей точность функционирования других алгоритмов, при использовании алгоритма «случайный лес» (т. е. этот алгоритм способен обнаруживать практически все атаки, присутствующие в нашем наборе данных) для надежного набора данных, содержащего современные атаки;

– сравнение алгоритмов машинного обучения;

– сравнение контролируемых алгоритмов с алгоритмом K-means, который является неконтролируемым алгоритмом, чтобы выяснить, какой из этих двух типов алгоритмов является наиболее надежным для обнаружения атак в сетях IoT.

2. Предлагаемая методика

В этом разделе представлена информация о разработанной IDS, ее функциях, выбранных моделях. Для разработки был выбран язык программирования Python, так как в нем есть развитые библиотеки обработки данных, обладающие множеством преимуществ.

Созданная программа работает в среде Python Jupyter Notebook и использует следующие пакеты [17]:

– pandas, numpy – поддержка массивов данных и операций с ними;

– mathplotlib, seaborn – библиотека визуализации данных;

– sklearn – платформа для анализа данных и машинного обучения.

Программа использует следующие модели машинного обучения:

– LogisticRegrade() алгоритма логистической регрессии;

– SVC() алгоритма SVC;

– RandomForestClassifier() алгоритма «случайный лес»;

– Kmeans() алгоритма K-means;

– KNeighboursClassifier() алгоритма KNN;

– GaussianNB() алгоритма Naive Bayes ;

– XGBClassifier() алгоритма XGBoost;

– AdaBoostClassifier() алгоритма AdaBoost;

– CatBoostClassifier() алгоритма CatBoost.

В ходе работы программы выполняются следующие шаги:

– подготовка предварительных данных как один из этапов работы модели;

– применение алгоритмов машинного обучения (логистическая регрессия, SVC, Random Forest, KNN, K-means, Naive Bayes, XGBoost, AdaBoost и CatBoost);

– определение наиболее значимых характеристик для каждого алгоритма;

– оценка качества модели по классификации: точность, ROC и AUC.

Табл. 1. Типы атак

Атаки	Количество записей	Описание атаки	Доля записей данного типа
Без атаки (Normal)	37 000	Естественные данные о транзакциях.	44,94%
Фаззерс (Fuzzerss)	6062	Попытка вызвать зависание программы или сети путем подачи на нее случайно сгенерированных данных.	7,36%
Анализ (Analysis)	677	Различные атаки порта, сканирование, проникновение спама и html-файлов.	0,82%
Бэкдоры (Backdoor)	583	Метод, при котором механизм скрытно обходит безопасность системы для получения доступа к компьютеру или его данным.	0,71%
Дос (Dos)	4 089	Злонамеренная попытка сделать сервер или сетевой ресурс недоступным для пользователей, обычно это временное прерывание или приостановка предоставления услуг хоста, подключенного к Интернету.	4,97%
Эксплойты (Exploits)	11 132	Злоумышленник знает о проблеме безопасности в операционной системе или части программного обеспечения и использует ее знание при эксплуатации уязвимости.	13,52%
Общий (Generic)	18 871	Техника работает против всех блоков шифра (с заданным блоком и ключом размер), без учета структуры блочного шифра.	22,92%
Разведка (Reconnaissance)	3 496	Содержит все удары, которые можно имитировать. Атаки, направленные на сбор информации.	4,25%
Шеллкод (Shellcode)	378	Маленький фрагмент программного кода использован как полезная нагрузка в эксплуатации его уязвимости.	0,46%
Черви (Worms)	44	Злоумышленник копирует себя, чтобы распространиться на другие компьютеры. Часто это используется в компьютерной сети для распространения себя и получения доступа к целевому компьютеру, полагаясь на сбои в системе безопасности.	0,05%
Всего	82 332	Естественные данные о транзакциях.	100%

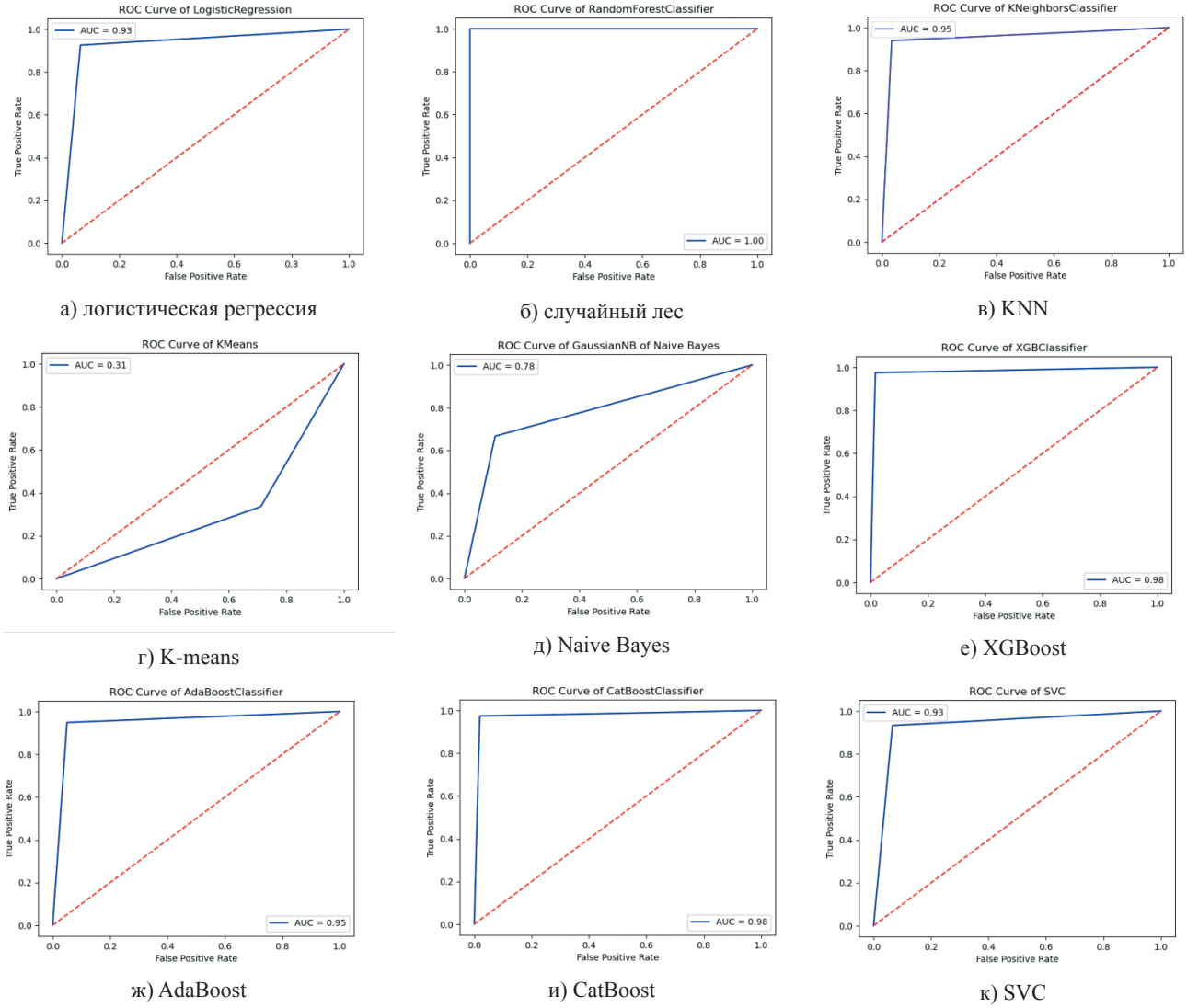
3. Сравнительные испытания и анализ

В этой части представлен исследуемый набор данных, выполнено тестирование выбранных моделей, дана оценка их производительности, осуществлен выбор лучшего алгоритма машинного обучения для обнаружения атак в сетях IoT.

Для оценки эффективности IDS необходим современный и полный набор данных, содержащий современные информацию о штатном функционировании системы и под действием атак [18]. Для проведения теста использован набор данных UNSW-NB15. UNSW-NB15 – это современный набор данных о реалистичных атаках и штатном функционировании вычислительных сетей [3]. Этот набор данных использовался в [19, 20] для статистических и оценочных целей путем сравнения пяти различных алгоритмов: кластеризация на основе деревьев решений, логистическая регрессия, наивный байесовский классификатор, приблизительные ближайшие соседи и ожидание-максимизация, чтобы измерить их производительность с точки зрения точности и уровень ложных тревог по сравнению с набором данных KDD99. Результаты оценки показали, что набор данных UNSW-NB15 можно считать более сложным и лучше отражающим современные атаки и обычный сетевой трафик, что делает его более подходящим для оценки методов атак существующей и предлагаемой системы обнаружения вторжений в сеть.

Этот набор данных содержит данные, соответствующие возможным реализациям 9 современных типов атак, и новые нормальные шаблоны трафика, а также 49 атрибутов, которые включают в себя потоки между хостами и проверку сетевых пакетов, чтобы различать нормальные или ненормальные наблюдения.

Набор данных разделен на набор для обучения (файл UNSW_NB15_training-set.csv содержит 82 332 записи разных типов) и набор для тестирования (файл UNSW_NB15_testing-set.csv содержит 175 341 запись разных типов). Подробности о представленных типах атак приведены в табл. 1. Количество данных об атаках в наборе для обучения превышает количество обычных транзакций: 45 332 по сравнению с 37 000.



а) логистическая регрессия

б) случайный лес

в) KNN

г) K-means

д) Naive Bayes

е) XGBoost

ж) AdaBoost

и) CatBoost

к) SVC

Рис. 3. Кривые ROC, полученные при использовании различных методов машинного обучения

Табл. 2. Сравнение значений метрик и оценок AUC различных алгоритмов

	Логистическая регрессия	SVC	KNN	Random Forest	K-means	Naive Bayes	XGBoost	AdaBoost	Catboost
Precision, %	93,02	93,36	95,19	97,69	31,95	79,53	97,87	94,97	97,73
Recall, %	92,98	93,34	95,11	97,68	31,46	76,84	97,86	94,96	95,72
F1-score, %	92,99	93,35	95,12	97,68	31,64	76,73	97,86	94,96	95,72
TPR	6 922	6 915	7 146	7 256	2 139	6 611	7 277	7 032	7 258
FPR	478	485	254	144	5261	789	123	368	142
FNR	678	611	551	238	6025	3025	229	462	234
TNR	8 389	8 456	8 516	8 829	3 042	6 042	8 838	8 605	8 833
Время обучения, с	101,41	160,76	10,93	6,76	0,64	0,16	12,59	10,15	32,56
AUC	0,93	0,93	0,95	0,98	0,31	0,78	0,98	0,95	0,98

Табл. 3. Сравнение результатов, полученных на обучающем и тестовом наборах данных

Обучающий набор данных									
	Логистическая регрессия	SVC	KNN	Random Forest	K-means	Naive Bayes	XGBoost	AdaBoost	Catboost
Precision, %	93,02	93,36	95,19	97,69	31,95	79,53	97,87	94,97	97,73
AUC	0,93	0,93	0,95	0,98	0,31	0,78	0,98	0,95	0,98
Длительность процесса тестирования, с	101,41	160,76	10,93	6,76	0,64	0,16	12,59	10,15	32,56
Тестовый набор данных									
Precision, %	93,87	93,93	94,71	95,94	40,81	84,41	95,76	94,26	95,78
AUC	0,91	0,90	0,94	0,95	0,33	0,84	0,95	0,92	0,95
Длительность процесса тестирования, с	188,94	547,31	35,90	18,21	1,04	0,22	28,62	23,55	57,62

Регулярные транзакции гораздо важнее любых атак. Общая атака является самой крупной среди атак, а атака «Черви» – самой маленькой. Для обучения моделей использованы данные из набора UNSW_NB15_training-set.csv, а для проверки надежности моделей – из набора UNSW_NB15_testing-set.csv.

В обучающем наборе использовано 16 467 записей, включая 9 067 записей об атаках и 7 400 записей без атак, а в тестовом наборе – 35 069 записей, включая 23 869 записей об атаках и 11 200 записей без атак.

Для каждого алгоритма рассчитаны значения его метрик и кривая ROC-AUC. Кривая ROC-AUC отображает соотношение истинно положительного результата (TPR) и уровня ложноположительного результата (FPR). AUC (площадь под кривой) будет использоваться в качестве единого значения для сравнения различных моделей для оценки их производительности. Оценки AUC варьируются от 0 до 1. Если оценка меньше 0,5, модель неэффективна и, следовательно, алгоритм не может быть успешно применен. Лучшим считается алгоритм с наибольшей AUC [21].

На рис. 3 показаны кривые ROC-AUC для рассмотренных алгоритмов.

В табл. 2 приведены значения метрик для различных методов машинного обучения. В этой таблице использованы следующие обозначения: Precision – точность распознания типа атак; Recall – доля правильно классифицированных объектов классов от числа всех объектов класса; F1-score – гармоническое значение Precision и Recall; FNR – ложноотрицательный результат; TNR – истинно-отрицательный результат.

Модели RandomForestClassifier, XGBClassifier, AdaBoostClassifier и CatBoostClassifier алгоритмов Random Forest, XGBoost, AdaBoost и CatBoost имеют самые высокие точности и самые высокие AUC.

Модель RandomForestClassifier быстрее трех последних, ее время выполнения составляет 6,76 секунды, а это означает, что этот алгоритм можно развернуть и применить с наибольшим успехом.

Результаты также показывают, что алгоритм K-means менее эффективен, чем другие рассмотренные алгоритмы обучения с учителем, хотя время его выполнения меньше, чем у Random Forest.

Это подтверждает, что алгоритмы контролируемого обучения более эффективны, чем алгоритмы неконтролируемого обучения, при обнаружении атак в сетях IoT.

При внедрении программных модулей:

– исходные данные были обработаны;

– реализованы прикладные алгоритмы машинного обучения;

– оценена точность обнаружения атак с помощью метрик машинного обучения;

– показана производительность каждого алгоритма с использованием кривой ROC;

– определено время выполнения алгоритмов;

– реализована программа в среде ноутбука Jupyter.

В табл. 3 сравниваются результаты работы алгоритмов на обучающем и тестовом наборах данных.

Работа на наборе тестовых данных дает очень хорошие результаты, аналогичные результатам, полученным на обучающем наборе данных. Это подтверждает качество обучения наших моделей.

Заключение

В этой статье проанализирована применимость таких методы машинного обучения, как логистическая регрессия, SVC, случайный лес, KNN, K-means, Naive Bayes, XGBoost, AdaBoost и CatBoost для обнаружения атак в сетях IoT. Точность и кривая ROC-AUC используются для оценки эффективности используемых моделей. Сравнивая 9 моделей используемых алгоритмов, выявлено, что алгоритм Random Forest имеет хорошую точность, самый высокий AUC и быстрое время выполнения, а это значит, что этот алгоритм можно развернуть и применить с наибольшим успехом. Будущая работа будет направлена на мультиклассификацию для различения типов атак.

Библиографический список

1. Arshad M.Z., Rahman H., Tariq J. et al. Digital Forensics Analysis of IoT Nodes using Machine Learning // Journal of Computing & Biomedical Informatics. 2022. Vol. 4. Issue 1. Pp. 1-12.

2. Overview of IoT threats in 2023. URL: https://securelist.com/iot-threat-report-2023/110644/ (дата обращения: 14.06.2024).

3. Chishakwe S., Ndlovu B.M., Dube S. et al. Intrusion Detection System for IoT environments using Machine Learning Techniques // 2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT). IEEE, 2022.

4. Hanif S., Ilyas T., Zeeshan M. Intrusion detection in IoT using artificial neural networks on UNSW-15 dataset // 2019 IEEE 16th international conference on smart cities: improving quality of life using ICT & IoT and AI (HONET-ICT). IEEE, 2019.

5. Liu Z., Thapa N., Shaver A. et al. Anomaly detection on iot network intrusion using machine learning // 2020 International conference on artificial intelligence, big data, computing and data communication systems (icABCD). IEEE, 2020.

6. Verma A., Ranga V. ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things // 2019 4th International conference on Internet of Things: Smart innovation and usages (IoT-SIU). IEEE, 2019.

7. Ikhwan S., Purwanto P., Rochim A.F. Comparison Analysis of Intrusion Detection using Deep Learning in IoT Networks // 2023 11th International Conference on Information and Communication Technology (ICoICT). IEEE, 2023.

8. Ahanger A.S., Khan S.M., Masoodi F.S. Intrusion Detection System for IoT Environment using Ensemble Approaches // 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2023.

9. Meng W. Intrusion detection in the era of IoT: Building trust via traffic filtering and sampling // Computer. 2018. Vol. 51. Issue 7. Pp. 36-43.

10. Wadate A.J., Deshpande S.P. Edge-Based Intrusion Detection using Machine Learning Over the IoT Network // 2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP). IEEE, 2023.

11. Alalade E.D. Intrusion detection system in smart home network using artificial immune system and extreme learning machine hybrid approach // 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). IEEE, 2020.

12. Кривошеин Б.Н., Покровский И.А. Понятия и критерии оценки технологической независимости и безопасности объектов критической информационной инфраструктуры // Безопасность информационных технологий. 2023. Т. 30. № 4. С. 39-60.

13. Алексеев В.М., Чичков С.Н. Защита информации в интеллектуальных транспортных системах управления городским транспортом // Надежность. 2022. № 22(3). С. 62-68. DOI: 10.21683/1729-26462022-22-3-62-68

14. Веселова В.А., Коломойцев В.С. Подход к обнаружению аномалий в самоподобном сетевом трафике // Надежность. 2023. № 23(2). С. 57-63. DOI: 10.21683/1729-2646-2023-23-2-57-63

15. Шубинский И.Б., Замышляев А.М., Проневич О.Б. и др. Применение методов машинного обучения для прогнозирования опасных отказов объектов железнодорожного пути // Надежность. 2020. № 20(2). С. 43-53. DOI: 10.21683/1729-2646-2020-20-2-43-53

16. Кулагин М.А., Сидоренко В.Г. Оценка экономической эффективности профилактических мероприятий по сокращению числа нарушений при управлении подвижным составом // Надежность. 2022/ № 22(4). С. 37-44. DOI: 10.21683/1729-2646-2022-22-4-37-44

17. Логинова Л.Н., Кулагин М.А. Применение Технологии Jupyter Notebook / Jupyter Hub для эффективного обучения в ВУЗах // Ректор ВУЗа. 2021. № 4. С. 32-37.

18. Belouch M., Hadaj S.E., Idhammad M. A two-stage classifier approach using reptree algorithm for network intrusion detection // International Journal of Advanced Computer Science and Applications. 2017. Vol. 8. No. 6. Pp. 389-394.

19. Nour M., Slay J. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set // Information Security Journal: A Global Perspective. 2016. Vol. 25. Nos. 1-3. Pp. 18-31.

20. Nour M., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) // 2015 military communications and information systems conference (MilCIS). IEEE, 2015.

21. Mitchell T.M. Machine Learning. New York: McGraw Hill, 1997. XVII, 414 p.

References

1. Arshad M.Z., Rahman H., Tariq J. et al. Digital Forensics Analysis of IoT Nodes using Machine Learning. *Journal of Computing & Biomedical Informatics* 2022;4(1):1-12.

2. Overview of IoT threats in 2023. (accessed 14.06.2024). Available at: https://securelist.com/iot-threat-report-2023/110644.

3. Chishakwe S., Ndlovu B.M., Dube S. et al. Intrusion Detection System for IoT environments using Machine Learning Techniques. In: 2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT). IEEE; 2022.

4. Hanif S., Ilyas T., Zeeshan M. Intrusion detection in IoT using artificial neural networks on UNSW-15 dataset. In: 2019 IEEE 16th international conference on smart cities: improving quality of life using ICT & IoT and AI (HONET-ICT). IEEE; 2019.

5. Liu Z., Thapa N., Shaver A. et al. Anomaly detection on iot network intrusion using machine learning. In: 2020 International conference on artificial intelligence, big data, computing and data communication systems (icABCD). IEEE; 2020.

6. Verma A., Ranga V. ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things. In: 2019 4th International conference on Internet of Things: Smart innovation and usages (IoT-SIU). IEEE; 2019.

7. Ikhwan S., Purwanto P., Rochim A.F. Comparison Analysis of Intrusion Detection using Deep Learning in IoT Networks. In: 2023 11th International Conference on Information and Communication Technology (ICoICT). IEEE; 2023.

8. Ahanger A.S., Khan S.M., Masoodi F.S. Intrusion Detection System for IoT Environment using Ensemble Approaches. In: 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE; 2023.

9. Meng W. Intrusion detection in the era of IoT: Building trust via traffic filtering and sampling. Computer 2018;51(7):36-43.

10. Wadate A.J., Deshpande S.P. Edge-Based Intrusion Detection using Machine Learning Over the IoT Network. In: 2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP). IEEE; 2023.

11. Alalade E.D. Intrusion detection system in smart home network using artificial immune system and extreme learning machine hybrid approach. In: 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). IEEE; 2020.

12. Krivoshein B.N., Pokrovsky I.A. Concepts and criteria for assessing the technological independence and security of critical information infrastructure facilities. *IT Security* 2023;30(4):39-60. (in Russ.)

13. Alekseev V.M., Chichkov S.N. Information security in intelligent mass transit management systems. *Dependability* 2022;22(3):62-68. DOI:10.21683/1729-26462022-2019-3-4-62-68 62- 68. (in Russ.)

14. Veselova V.A., Kolomoytsev V.S. An approach to detecting anomalies in a self-similar network traffic. *Dependability* 2023;23(2):57-63. DOI: 10.1007/1729-2646-2023-23-2- 57- 63. (in Russ.)

15. Shubinsky I.B., Zamyshliaev A.M., Pronevich O.B. et al Application of machine learning methods for predicting hazardous failures of railway track assets. *Dependability* 2020;2:45-53. DOI: 10.21683/1729-2646-2020-20-2-43-53.

16. Kulagin M.A., Sidorenko V.G. Evaluating the economic efficiency of preventive measures aimed at reducing the number of train control violations. *Dependability* 2022;22(4):37-44. DOI: 10.21683/1729-2646-201322-4-37. (in Russ.)

17. Loginova L.N., Kulagin M.A. Application of Jupiter Notebook Technology. Jupiter Hub for effective education in universities. *Rektor VUZa* 2021;4:32-37. (in Russ.)

18. Belouch M., Hadaj S.E., Idhammad M. A two-stage classifier approach using reptree algorithm for network intrusion detection. *International Journal of Advanced Computer Science and Applications* 2017;8(6):389-394.

19. Nour M., Slay J. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective* 2016;25(1-3):18-31.

20. Nour M., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: 2015 military communications and information systems conference (MilCIS). IEEE; 2015.

21. Mitchell T.M. Machine Learning. New York: McGraw Hill; 1997.

Сведения об авторах

Нианг Папа Малик – аспирант РУТ (МИИТ), ул. Образцова, д.9, стр.9, Москва, Российская Федерация, 127994, e-mail: malickdiarra30@gmail.com

Сидоренко Валентина Геннадьевна – доктор технических наук; профессор; профессор кафедры «Управление и защита информации» РУТ (МИИТ), ул. Образцова, д.9, стр.9, Москва, Российская Федерация, 127994, e-mail: valenfalk@mail.ru

About the authors

Niang Papa Malick, Postgraduate Student, RUT (MIIT), 9b9 Obrazcova St., Moscow, 127994, Russian Federation, e-mail: malickdiarra30@gmail.com

Valentina G. Sidorenko, Doctor of Engineering, Professor, Chair Professor, Department of Management and Protection of Information, RUT (MIIT), 9b9 Obrazcova St., Moscow, 127994, Russian Federation, e-mail: valenfalk@mail.ru.

Вклад авторов в статью

Нианг П.М. Проведен обзор литературы по теме исследования. Выбраны и реализованы девять алгоритмов машинного обучения для определения лучшего для обнаружения вторжений в устройства IoT путем бинарной классификации.

Сидоренко В.Г. Анализ полученных результатов.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.