УДК 004

Гаврилова А.А., Мусина К.Ш.
Уфимский университет науки и технологий, Уфа

Научный руководитель:
Юнусова Д.С.
Уфимский университет науки и технологий, Уфа

ОПРЕДЕЛЕНИЕ И ОБНАРУЖЕНИЕ DDoS-АТАК С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В статье представлено исследование применения различных методов машинного обучения для обнаружения DDoS-атак на основе анализа сетевого трафика. Проводится сравнительный анализ производительности моделей Random Forest, Bagging, AdaBoost, XGBoost, LightGBM, CatBoost с использованием метрик Accuracy, Precision, Recall, F1 и ROC-AUC. Целью работы является определение наиболее эффективного метода для раннего обнаружения и предотвращения DDoS-атак.

Ключевые слова: DDoS-атаки, кибербезопасность, машинное обучение, анализ сетевого трафика, предотвращение атак, Accuracy, Precision, Recall, F1, ROC-AUC.

Обеспечение надежной защиты сетевой инфраструктуры критически важно для предприятий различного масштаба ввиду возрастающего риска возникновения DDoS-атак. Классические способы противодействия данным угрозам зачастую недостаточно эффективны, что обусловлено постоянным усложнением и увеличением объемов атак. Один из перспективных путей улучшения ситуационной осведомленности и раннего реагирования на подобные инциденты заключается в применении методов машинного обучения, обладающих возможностью автоматически анализировать сетевой трафик и выделять характерные паттерны поведения, присущие DDoS-атаке.

Статья направлена на решение задачи определения и выявления DDoS-атак с использованием методов машинного обучения. Основой исследования послужил открытый набор данных, состоящий из записей сетевого трафика, собранных в процессе наблюдения за деятельностью злоумышленников. Общий объем набора составляет 1048444 записей, каждая из которых описывает отдельные события сетевого взаимодействия, включая протоколы верхних уровней и транспортных слоев, IP-адреса отправителей и получателей, порты соединений, размеры передаваемых пакетов и интервал времени между передачами.

Предварительно была проведена обработка данных. Реализована процедура исключения полей с IP-адресами отправителя и получателя, поскольку эти данные не оказывают прямого влияния на природу сетевого

трафика и способствуют увеличению вычислительной нагрузки без увеличения информативности. Отсутствующие значения были заполнены средним значением соответствующего столбца, используя метод mean imputation. Средняя величина дает нейтральную оценку отсутствующему значению, минимизируя возможные смещения в дальнейшей оценке.

Категориальные признаки, такие как протоколы высших и нижних уровней, были переведены в числовую форму с помощью OneHotEncode. Данная техника преобразования позволила сохранить всю необходимую информацию о категориях и подготовить данные для ввода в модели машинного обучения.

Выбор конкретных методов объясняется их популярностью и доказанной эффективностью в аналогичных задачах. Были использованы следующие алгоритмы:

– Random forest: базируется на создании ансамбля решающих деревьев, способных одновременно решать задачи классификации и регрессии, устойчив к переобучению и работает быстро даже с большими объемами данных;

– Bagging classifier: является простым и эффективным способом повышения точности за счет уменьшения вариативности базовой модели;

– Adaboost: алгоритм адаптивного бустинга, последовательно улучшая точность за счет коррекции ошибок предыдущих итераций;

– Xgboost: оптимизированная реализация градиентного бустинга, популярная в соревнованиях по анализу данных благодаря высокой скорости и качеству работы;

– Lightgbm: еще одна версия градиентного бустинга, известная своим быстрым временем обучения и низкими требованиями к ресурсам;

– CatBoost: выделяется высоким качеством работы с категориальными признаками и отсутствием необходимости специальной предобработки данных.

Данные были разделены на обучающую (70 %) и тестовую (30 %) выборки с использованием параметра stratify=y в функции train_test_split(). Данный параметр гарантирует сохранение первоначального распределения классов в обеих выборках, что предотвращает смещение результатов. Такое разделение оптимально, так как оно обеспечивает достаточную долю данных для качественного обучения модели и позволяет провести независимую проверку ее способностей к обобщению на незнакомые образцы. Модели были обучены на обучающей выборке, а их производительность оценивалась с использованием метрик точности (accuracy), точности позитивных предсказаний (precision), чувствительности (recall), меры F1 и показателя ROC-AUC.

В табл. 1 представлены результаты оценки производительности методов машинного обучения при обнаружении DDoS-атак на основе анализа сетевого трафика, где сравнивались значения Accuracy, Precision, Recall, F1 и ROC-AUC.

Таблица 1

Оценка производительности методов машинного обучения

	Accuracy	Precision	Recall	F1	ROC-AUC
CatBoost	0,998073	0,999101	0,996765	0,997932	0,999993
XGBoost	0,998022	0,998548	0,997209	0,997878	0,999993
LightGBM	0,998006	0,999051	0,996673	0,997860	0,999993
Bagging Classifier	0,998026	0,998624	0,997142	0,997882	0,999993
Random Forest	0,998026	0,998624	0,997142	0,997882	0,999993
AdaBoost	0,994781	0,988944	0,999992	0,994437	0,999956

По итогам проведенного исследования выяснилось, что лучшие результаты обеспечивали ансамблевые методы CatBoost и XGBoost, стабильно демонстрируя самую высокую общую точность и лучшую обобщающую способность. Оба метода достигли показателя точности более 99 %, что фактически соответствует уровню практической полезности для реальных систем мониторинга и защиты. Менее выраженные различия наблюдались между другими моделями, такими как Random Forest, Bagging Classifier и LightGBM, которые также дали хорошие результаты, но незначительно уступали двум лидирующим моделям.

Важно подчеркнуть, что такая высокая точность достигается благодаря уникальной архитектуре и оптимизации данных моделей, что позволяет учитывать широкий спектр признаков и успешно выявлять даже небольшие отклонения в поведении сетевого трафика, ассоциированные с DDoS-атаками.

Анализируя результаты моделей, можно выделить несколько ключевых признаков, которые могут указывать на наличие DDoS-атаки:

– частота пакетов (Packets/Time): в случае DDoS-атаки наблюдается резкое увеличение количества пакетов в единицу времени, что создает аномальную нагрузку на сеть;

– протоколы (Highest Layer, Transport Layer): использование специфичных протоколов, таких как ARP или ICMP, может свидетельствовать о попытках перегрузки сети;

– длина пакета (Packet Length): необычные размеры пакетов, особенно если они значительно отличаются от типичного сетевого трафика, могут сигнализировать о подозрительной активности.

Эти признаки играют ключевую роль в определении наличия DDoS-атаки, позволяя моделям машинного обучения эффективно распознавать аномалии в сетевом трафике.

Таким образом, данная работа иллюстрирует успешность применения методов машинного обучения для разработки эффективных систем выявления DDoS-атак. Рассмотренный подход открывает перспективы для

интеграции интеллектуальных аналитических модулей в корпоративные системы безопасности, повышая устойчивость и защищенность инфраструктуры от растущих угроз.

Список использованных источников:

1. Иванникова В.П. Бинарная классификация компьютерных атак на примере базы данных UNSW-NB15 / В.П. Иванникова, О.И. Шелухин // Телекоммуникации и информационные технологии. 2020. Т. 7, № 1. С. 10–18.

2. Колосов И.А. Классический подход решения задачи бинарной классификации / И.А. Колосов, К.Т. Джахангирова // Научный аспект. 2024. Т. 3, № 6. С. 278–283.

Gavrilova A.A., Musina K.S.
Ufa University of Science and Technology, Ufa, Russia

Scientific supervisor:
Yunusova D.S.
Ufa University of Science and Technology, Ufa, Russia

IDENTIFICATION AND DETECTION OF DDoS ATTACKS USING MACHINE LEARNING METHODS

Abstract. The article presents a study of the use of various machine learning methods for detecting DDoS attacks based on network traffic analysis. A comparative analysis of the performance of the Random Forest, Bagging, AdaBoost, XGBoost, LightGBM, and CatBoost models is carried out using Accuracy, Precision, Recall, F1, and ROC-AUC metrics. The aim of the work is to determine the most effective method for early detection and prevention of DDoS attacks.

Keywords: DDoS attacks, cybersecurity, Machine learning, network traffic analysis, attack prevention, Accuracy, Precision, Recall, F1, ROC-AUC.