УДК 004.8

Зыков А.И.
Уфимский университет науки и технологий, Уфа

Научный руководитель:
Шафиков М.Р.
Уфимский университет науки и технологий, Уфа

ПРИМЕНЕНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЗАЩИТЫ КРИТИЧЕСКОЙ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ ОТ СЕТЕВЫХ АТАК

Аннотация. В статье рассматривается применение современных технологий искусственного интеллекта и алгоритмов машинного обучения

для обнаружения сетевых кибератак на объектах критической информационной инфраструктуры РФ. Фокус в статье сделан на написании программного кода для анализа вредоносного сетевого трафика с помощью предопределенных меток и машинного обучения без данных меток.

Ключевые слова: искусственный интеллект, машинное обучение, информационная безопасность, защита информации, кибератаки.

Обеспечение защищенности конфиденциальной информации – одна из приоритетных задач современного общества Российской Федерации. Это обусловлено стремительным развитием цифровых технологий и увеличением объемов сетевого трафика, с ростом которых неумолимо растет число кибератак. Большинство из них направлено на критическую информационную инфраструктуру (КИИ). Объекты КИИ представляют собой ключевые компоненты информационной экосистемы государства и бизнеса, атаки на которые могут привести к серьезным последствиям для государства.

Согласно данным TAdviser, с каждым годом наблюдается рост числа кибератак на объекты КИИ РФ. Если в 2020 г. число атак составляло около 120000, то к 2024 г. это число выросло более, чем в 5 раз. Это означает, что, в среднем, ежемесячно субъектам КИИ РФ приходится отражать более 20000 атак. Результатом анализа, проведенным компанией RED Security, стало то, что самой атакуемой отраслью в сфере КИИ в 2024 г. стала промышленность: ТЭК, энергетика, оборонная промышленность и др. На их долю пришлось около 1/3 от общего числа кибератак на КИИ [2]. За сферой промышленности следует область ИТ. На нее пришлось 20 % от всех атак из-за возросшей популярности кибератак на российские компании через их ИТ-подрядчиков – вендоров, интеграторов и т. п. Третьей по популярности атак киберпреступников считается финансовый сектор. В 2024 г. на банковские и другие финансовые предприятия было совершено 17 % от всего числа целенаправленных атак.

В контексте обеспечения сетевой безопасности искусственный интеллект и алгоритмы машинного обучения позволяют анализировать сетевой трафик в режиме реального времени, выявлять аномалии, прогнозировать угрозы и автоматически классифицировать кибератаки с высокой точностью. Согласно статистическим данным экспертов от Anti-Malware, более трети от всех ожидаемых инноваций в сфере информационной безопасности предписывают именно технологиям искусственного интеллекта и алгоритмам машинного обучения.

Наиболее широкое применение ИИ и МО получили в сфере сетевой защиты. Наиболее подходящим средством обеспечения безопасности данных от сетевых вторжений является система обнаружения вторжений с алгоритмами машинного обучения (ML IDS). При применении искусственного интеллекта и машинного обучения на этапе проектирования средств и систем готовый продукт должен отвечать требованиям как

ГОСТ Р 71476-2024, так и ГОСТ Р 71539-2024. Особенностью построения и применения таких систем является подготовка отдельной базы данных/знаний для обучения и настройки моделей, а также необходимость в верификации и непрерывной валидации – тестировании модели перед и после развертывания соответственно [1]. В ML IDS валидация модели машинного обучения необходима для обеспечения подтверждения их эффективности и точности в обнаружении угроз.

Проведем тестирование алгоритма машинного обучения, применяемый в ML IDS. Для этого смоделируем сетевую кибератаку на объект КИИ – автоматизированную систему управления технологическими процессами. Для этого возьмем готовый датасет CIS-IDS-2017, включающий нормальный трафик и различные виды атак, имитирующие реальные сценарии сетевой безопасности: DoS- и DDoS-атаки, Port Scan, Brute Force, Web Attacks, Infiltration, Botnet Activity, Heartbleed.

Для проведения программного анализа по тем меткам, которые присутствуют в датасете, отбросим трафик, соответствующий нормальной человеческой активности (BENIGN).

Согласно проведенному анализу (рис. 1), выделим следующее: в датасете присутствует 8 видов различных кибератак, их суммарное количество 425878, более 3/4 от всех атак пришлось на DoS- и DDoS-атаки – 45,5 % и 30 % соответственно, а наименьшая доля – на «Infiltration» и «Heartbleed».

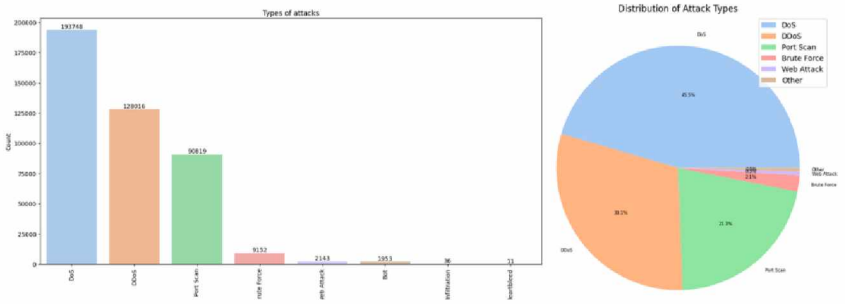

Рис. 1. Количество сетевых атак по видам (слева) и их процентное соотношение (справа) на основании программного анализа датасета

Получив исходные данные, можно приступить к обучению алгоритма для предсказания сетевых атак. Для обучения был выбран алгоритм машинного обучения – метод случайного леса (RandomForest), создающий случайные леса для решения задач классификации и регрессии [3]. Для того, чтобы обучить модель детектировать атаки, необходимо подготовить данные – удалить метку «Тип атаки», чтобы алгоритм МО обучался без подсказок, а также удалить BENIGN, разделить данные на обучающую и тестовую выборку в соотношении 80 % к 20 %, обучить алгоритм и провести оценку обученности алгоритма на тестовой выборке. Код программы по обучению модели показан на рис. 2.

```
filtered_data = data.loc[data['Attack Type'] != 'BENIGN']

features = filtered_data.drop(['Attack Type'], axis=1)
labels = filtered_data['Attack Type']

X_train, X_test, y_train, y_test = train_test_split( *arrays: features, labels, test_size=0.2, random_state=42)

model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

predictions = model.predict(X_test)
accuracy = accuracy_score(y_test, predictions)
print(f'Accuracy: {accuracy:.4f}')

print(classification_report(y_test, predictions))

plt.figure(figsize=(10, 6))
conf_matrix = confusion_matrix(y_test, predictions)
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=model.classes_, yticklabels=model.classes_)
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.title('Confusion Matrix')
plt.show()
```

Рис. 2. Фрагмент программного кода для обучения модели RandomForest по предсказанию сетевых атак датасета CIS-IDS-2017

Точность детектирования классов сетевых атак – 0,9998 из 1. Как видно из отчета (рис. 3), модель ошиблась в распознавании веб-атак, однако показатель правильных ответов также высокий – 0,99 из 1.

Accuracy: 0.9998

	precision	recall	f1-score	support
Bot	1.00	1.00	1.00	430
Brute Force	1.00	1.00	1.00	1856
DDoS	1.00	1.00	1.00	25653
DoS	1.00	1.00	1.00	38624
Heartbleed	1.00	1.00	1.00	3
Infiltration	1.00	1.00	1.00	6
Port Scan	1.00	1.00	1.00	18190
Web Attack	0.99	1.00	0.99	414
accuracy			1.00	85176
macro avg	1.00	1.00	1.00	85176
weighted avg	1.00	1.00	1.00	85176

Рис. 3. Оценка обученности модели RandomForest

Результатом сравнения графиков результатов программного анализа и алгоритма машинного обучения (рис. 4) является: общее количество сетевых атак пропорционально меньше при тесте ИИ-модели и составляет 20 % от всего датасета – 85176 и 425878 атак, в долевом соотношении наблюдается практически полное совпадение с результатами программного анализа, градация сетевых атак по частоте встречаемости также совпадает.

В результате проведенного анализа делаем вывод, что алгоритмы МО и ИИ способны с большой точностью выявлять атаки. Искусственный интеллект и машинное обучение – это передовые технологии, внедрение которых в системы безопасности КИИ является важным для защиты объектов. А активное развитие и применение ИИ и МО в области ИБ – необходимость современности, чтобы справляться с постоянно растущим количеством киберугроз и уязвимостей.

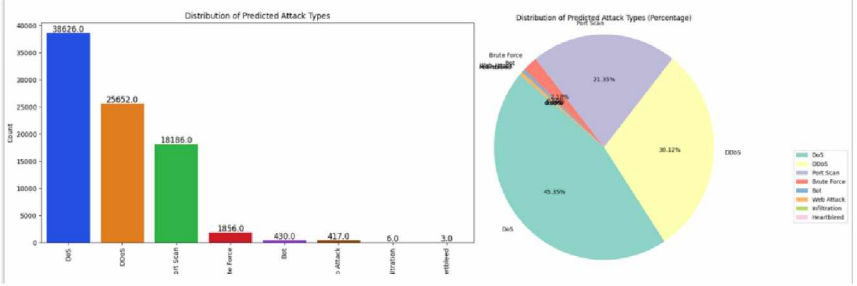

Рис. 4. Количество сетевых атак по видам (слева) и их процентное соотношение (справа) на основании предсказания модели RandomForest

Список использованных источников:

1. ГОСТ Р 71539-2024 (ИСО/МЭК 5338:2023). Национальный стандарт Российской Федерации. Искусственный интеллект. Процессы жизненного цикла системы искусственного интеллекта // КонсультантПлюс: [сайт]. URL: https://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=OTN&n=43150#OU7AUhUSG1BHGtM81 (дата обращения: 30.03.2025).
2. RED Security SOC: хакеры усилили давление на критическую информационную инфраструктуру России // Компания «RED Security»: [сайт]. URL: https://redsecurity.ru/news/red-security-soc-khakery-usilili-davlenie-na-kriticheskuyu-informatsionnuyu-infrastrukturu-rossii (дата обращения: 30.03.2025).
3. Баланов А.Н. Машинное обучение и искусственный интеллект: учебное пособие для вузов [Текст] / А.Н. Баланов. 2-е изд. СПб.: Лань, 2025. 172 с.

Zykov A.I.
Ufa University of Science and Technology, Ufa

Scientific supervisor:
Shafikov M.R.
Ufa University of Science and Technology, Ufa

USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING METHODS TO PROTECT CRITICAL INFORMATION INFRASTRUCTURE FROM NETWORK ATTACKS

Abstract. In the article it is discussed using of modern artificial intelligence technologies and machine learning algorithms to detect network cyberattacks on critical information infrastructure objects of Russian Federation. The focus of the article is on writing software code to analyze malicious network traffic using predefined labels and a machine learning algorithm without these labels.

Keywords: artificial intelligence, machine learning, information security, information protection, cyberattacks.