ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ И ИНСТРУМЕНТОВ ОЦЕНКИ УРОВНЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Сардаров Ягуб
Алиев Камил
Азербайджанский государственный университет нефти и промышленности

Аннотация
Данное исследование рассматривает статистические методы и инструменты для оценки уровня информационной безопасности. Описаны основные концепции и подходы к оценке рисков, методы анализа уязвимостей и инцидентов безопасности. Рассмотрены современные статистические модели для прогнозирования угроз и оценки вероятности их реализации, включая методы машинного обучения, такие как классификация, регрессия, кластерный анализ и обнаружение аномалий. Исследование также охватывает инструменты для анализа данных о безопасности, включая аналитику больших данных. Описаны практические примеры применения этих методов и их преимущества. Полученные результаты могут быть полезны для специалистов в области информационной безопасности при выборе оптимальных методов и инструментов для защиты информации, также снижения рисков. Описанные подходы и модели могут помочь в разработке новых стратегий обеспечения безопасности и защите от современных киберугроз.
Ключевые слова: информационная безопасность, машинное обучение, анализ данные.
Введение. В современном информационном обществе данные являются ценным активом, и их безопасность - приоритетная задача. Угроза кибератак требует эффективных методов защиты информации [1, 2]. Одной из новых возможностей является использование методов машинного обучения. Машинное обучение автоматизирует обнаружение угроз, анализ уязвимостей и предсказание инцидентов безопасности. Эти методы обрабатывают большие объемы данных, выявляя скрытые закономерности и аномалии. Исследование фокусируется на применении методов машинного обучения для оценки информационной безопасности. Рассматриваются различные модели для обнаружения угроз, анализа уязвимостей и предсказания инцидентов, а также их использование для создания систем защиты данных и предотвращения кибератак. Особое внимание уделяется методам классификации, кластеризации, обнаружения аномалий и регрессионного анализа. Эти методы помогают разрабатывать стратегии управления рисками и улучшать защитные меры. Цель исследования - предоставить обзор текущих тенденций и перспектив использования методов машинного обучения для повышения уровня информационной безопасности, чтобы специалисты могли найти оптимальные решения для защиты информации.
Методы. Основные методы машинного обучения следующие.
Классификация. Это включает обзор алгоритмов классификации, их принцип работы и применение в оценке уровня информационной безопасности [3]. Классификация в машинном обучении относится к задаче разделения данных на категории или классы на основе их характеристик. Например, метод опорных векторов (SVM) может быть применен для определения типа кибератаки на основе характеристик сетевого трафика или других данных. Наивный Байесовский классификатор, с другой стороны, основан на теореме Байеса и может использоваться для определения категории уязвимости на основе известных признаков уязвимостей.
Кластеризация. Здесь рассматривается применение алгоритмов кластеризации для выявления групп подобных событий или данных без заранее известных категорий. Это позволяет обнаруживать аномалии и потенциальные угрозы, основываясь на сходстве между данными. Например, кластеризация может помочь выявить подозрительные схемы поведения

пользователей или необычные сетевые активности, которые могут указывать на наличие вредоносного поведения [1].

Регрессия. В этом разделе рассматривается использование методов регрессии для прогнозирования вероятности возникновения инцидентов безопасности или оценки уровня риска. Регрессия в машинном обучении используется для построения моделей, которые предсказывают значения на основе других переменных. Например, регрессионные модели могут использоваться для оценки вероятности успешной атаки на основе известных параметров атакующего и защищающегося компьютеров, а также степени уязвимости системы на основе ее характеристик. Теперь рассмотрим использованию обучения на основе данных [5].

Обнаружение аномалий. При этом методы обнаружения аномалий позволяют выявлять необычные или подозрительные события в потоке данных. Изоляционный лес и Одноклассовый метод опорных векторов - два широко используемых метода для этой цели. Изоляционный лес работает путем создания леса из решающих деревьев, разбивая данные на случайные подмножества, что позволяет быстро выявлять аномалии, так как они требуют меньше разбиений. Он эффективно обрабатывает данные с большим числом измерений и способен выделять аномалии даже в больших наборах данных. Одноклассовый метод опорных векторов строит гиперплоскость, которая отделяет нормальные данные от аномалий, что позволяет обнаруживать необычные паттерны. Этот метод подходит для данных высокой размерности и хорошо работает с выбросами в данных [4].

Код для примера:

```
import numpy as np
import pandas as pd
from sklearn.ensemble import IsolationForest
from sklearn.svm import OneClassSVM
from sklearn.metrics import classification_report, accuracy_score
import matplotlib.pyplot as plt
def generate_network_traffic(n_samples=1000, n_features=5, anomaly_fraction=0.05):
    np.random.seed(42)
    X_normal = np.random.normal(0, 1, size=(int(n_samples * (1 - anomaly_fraction)), n_features))
    X_anomaly = np.random.normal(5, 1, size=(int(n_samples * anomaly_fraction), n_features))
    X = np.vstack((X_normal, X_anomaly))
    y = np.hstack((np.zeros(int(n_samples * (1 - anomaly_fraction))), np.ones(int(n_samples * anomaly_fraction))))
    return X, y
X, y = generate_network_traffic()
isolation_forest = IsolationForest(contamination=0.05, random_state=42)
y_pred_if = isolation_forest.fit_predict(X)
y_pred_if = np.where(y_pred_if == 1, 0, 1)
svm = OneClassSVM(nu=0.05, kernel="rbf", gamma=0.1)
y_pred_svm = svm.fit_predict(X)
y_pred_svm = np.where(y_pred_svm == 1, 0, 1)
def print_results(y_true, y_pred, method_name):
    print(f"Результаты по {method_name}:")
    print(classification_report(y_true, y_pred))
    print(f"Accuracy: {accuracy_score(y_true, y_pred):.2f}")
    print("\n")
print_results(y, y_pred_if, "Изоляционный лес")
print_results(y, y_pred_svm, "Одноклассовый метод опорных векторов")
```

```
def visualize_results(X, y_true, y_pred, method_name):
    plt.figure(figsize=(12, 6))
    plt.scatter(X[:, 0], X[:, 1], c=y_pred, cmap='coolwarm', label='Предсказанные Аномалии')
    plt.scatter(X[:, 0], X[:, 1], c=y_true, cmap='viridis', alpha=0.5, marker='x', label='Настоящие Аномалии')
    plt.title(f"{method_name} - Обнаружение аномалий")
    plt.xlabel("Свойство 1")
    plt.ylabel("Свойство 2")
    plt.legend()
    plt.show()
visualize_results(X, y, y_pred_if, "Изоляционный лес")
visualize_results(X, y, y_pred_svm, "Одноклассовый метод опорных векторов")
```

Результат:

Результаты по Изоляционный лес:

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	950
1.0	0.98	0.98	0.98	50
accuracy			1.00	1000
macro avg	0.99	0.99	0.99	1000
weighted avg	1.00	1.00	1.00	1000

Accuracy: 1.00

Результаты по Одноклассовый метод опорных векторов:

	precision	recall	f1-score	support
0.0	0.97	0.97	0.97	950
1.0	0.35	0.34	0.35	50
accuracy			0.94	1000
macro avg	0.66	0.65	0.66	1000
weighted avg	0.93	0.94	0.94	1000

Accuracy: 0.94

Наглядная демонстрация результатов по предскаханию аномалий с использованием методов как Изоляционный лес и Одноклассовый метод опорных векторов также представлены на рисунках 1, 2 соответственно.

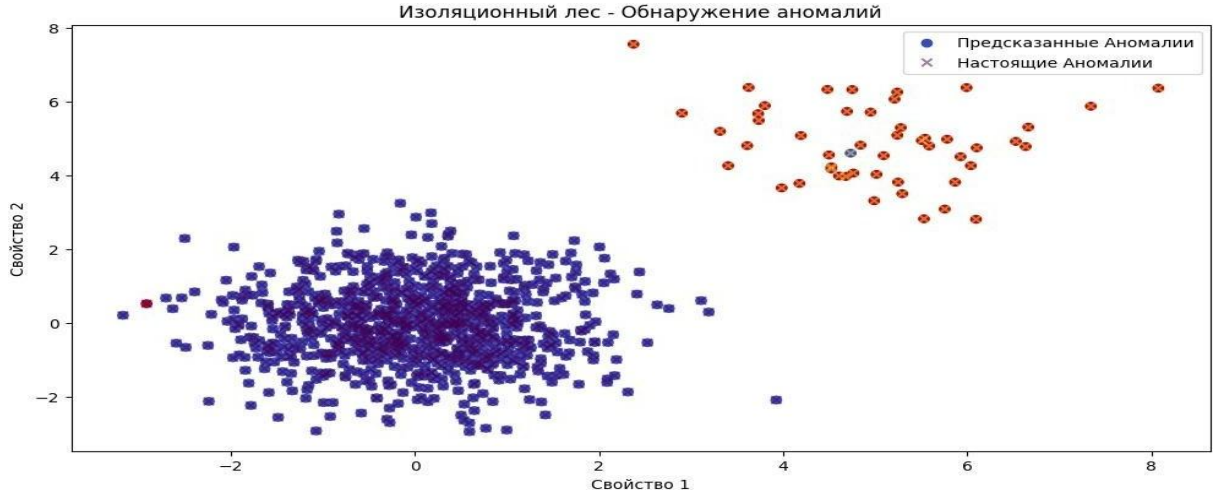

Рис. 1. Визуализация данных с соответствующими предсказаниями по методу Изоляционный лес

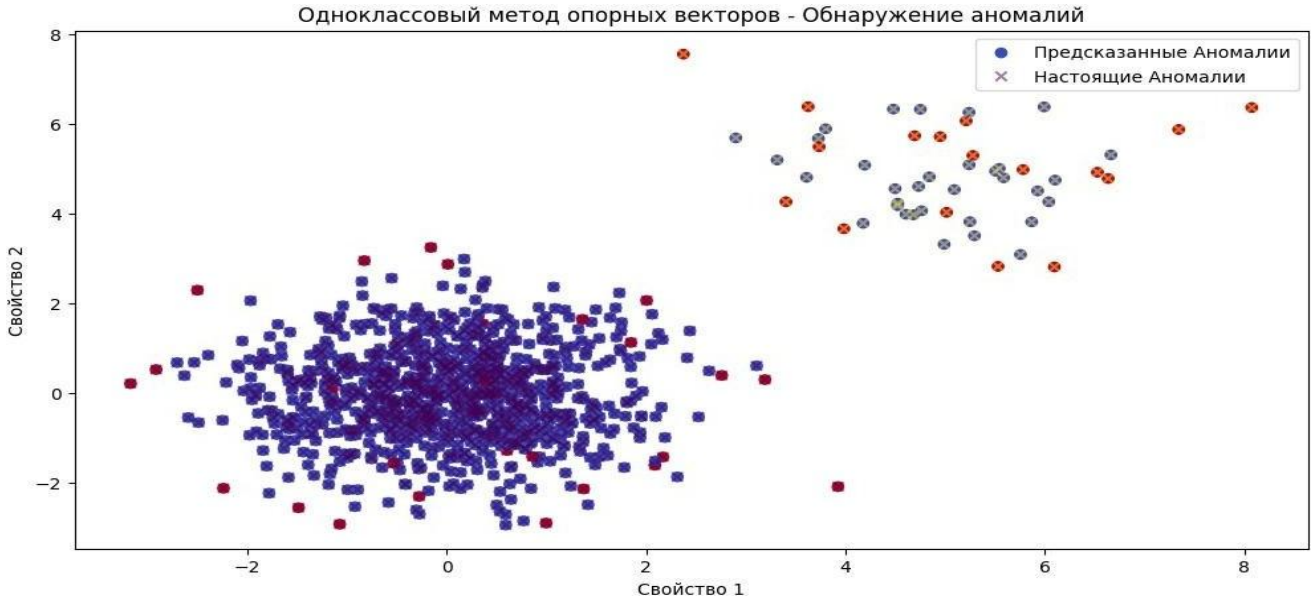

Рис. 2. Визуализация данных с соответствующими предсказаниями по Одноклассовому методу опорных векторов

Функция generate_network_traffic: Эта функция создает набор данных с нормальными и аномальными точками.

X_normal: генерирует нормальные данные из нормального распределения с параметрами (0, 1).

X_anomaly: генерирует аномальные данные из нормального распределения с параметрами (5, 1).X: объединяет нормальные и аномальные данные в один массив.

y: создает метки для нормальных данных (0) и аномальных данных (1).

Обнаружение аномалий с использованием Изоляционный лес

Изоляционный лес: Модель Изоляционный лес обучается на наборе данных и используется для предсказания аномалий.

y_pred_if: результат предсказания модели. Метки изменяются на 0 для нормальных данных и 1 для аномалий.

Обнаружение аномалий с использованием Одноклассового метода опорных векторов. Одноклассовый метода опорных векторов: Модель Одноклассового метода опорных векторов обучается и используется аналогично для обнаружения аномалий.y_pred_svm: результат предсказания модели. Метки изменяются на 0 для нормальных данных и 1 для аномалий. Функция print_results: Эта функция выводит отчет о классификации и точность для каждого метода.classification_report: показывает точность, полноту и F1-оценку.

accuracy_score: показывает точность предсказания.

Функция visualize_results (опционально):Визуализирует первые два признака набора данных, показывая истинные и предсказанные аномалии для каждого метода.

plt.scatter: используется для создания графика рассеяния данных, где точки окрашены в зависимости от меток (истинные или предсказанные аномалии).

Прогнозирование инцидентов безопасности. Рассматривается применение алгоритмов классификации и регрессии для предсказания вероятности возникновения конкретных инцидентов безопасности [4]. Модели машинного обучения могут использоваться для определения, какие угрозы наиболее вероятно возникнут в будущем, и принятия соответствующих мер по их предотвращению [5]. Например, модель классификации может предсказывать типы атак на основе анализа предшествующих инцидентов, а модель регрессии

может оценивать вероятность успешного вторжения на основе характеристик сети и обнаруженных уязвимостей. Результаты такого прогнозирования могут помочь организациям адаптировать свои стратегии защиты и реагирования на угрозы информационной безопасности.

Обработка больших данных. Обработка больших данных играет ключевую роль в обеспечении информационной безопасности организаций, позволяя обнаруживать угрозы и предотвращать атаки в режиме реального времени. Рассмотрим основные технологии и методы обработки больших данных, применяемые в этой области [2].

Технологии потоковой обработки данных. Важным аспектом в обработке данных о безопасности является возможность обрабатывать потоки данных в реальном времени. Для этого часто применяются технологии потоковой обработки, такие как Apache Kafka и Apache Flink. Apache Kafka обеспечивает эффективную передачу и обработку потоков данных из различных источников, что позволяет оперативно реагировать на угрозы. Apache Flink, в свою очередь, предоставляет возможности для высокоэффективной обработки потоков данных, включая выполнение сложных вычислений и анализ данных в реальном времени. Эти технологии позволяют организациям оперативно реагировать на угрозы и предотвращать атаки до их реализации [4].

Методы распределенных вычислений. Обработка крупномасштабных данных о безопасности также требует эффективных методов распределенных вычислений. Apache Spark является одним из основных инструментов в этой области. Он позволяет обрабатывать огромные объемы данных и извлекать из них ценную информацию о потенциальных угрозах. Apache Spark предоставляет API для работы с данными как в памяти, так и на диске, что делает его идеальным выбором для обработки больших данных [3]. Он также поддерживает параллельное выполнение операций, что позволяет обрабатывать данные быстро и эффективно.

Использование этих технологий и методов обработки больших данных позволяет организациям эффективно анализировать данные о безопасности, выявлять аномалии и предотвращать возможные угрозы. Они играют важную роль в обеспечении информационной безопасности и защите данных от кибератак.

Архитектуры нейронных сетей. Глубокое обучение на основе нейронных сетей применяется для анализа данных о безопасности с целью выявления угроз и защиты информации [2]. Сверточные нейронные сети (CNN) эффективно работают с изображениями, позволяя обнаруживать сложные закономерности в визуальных данных. Например, они могут использоваться для распознавания образов на видеозаписях или анализа содержимого изображений с камер видеонаблюдения, что помогает в обнаружении необычных или подозрительных действий. Рекуррентные нейронные сети (RNN), в свою очередь, применяются для анализа последовательных данных, таких как тексты или временные ряды. Они способны выявлять сложные зависимости в последовательных данных, что может быть полезно для выявления угроз в текстовых сообщениях или обнаружения аномалий во временных рядах событий [4, 5].

Перенос обучения и генеративные модели. Перенос обучения - это метод, который позволяет использовать знания, полученные из одной задачи, для улучшения результатов в другой задаче. Например, модель, обученная на крупномасштабных данных о вредоносном ПО, может быть адаптирована для анализа новых угроз без необходимости обучения на новых данных с нуля. Это позволяет сократить время и затраты на обучение новых моделей. Генеративные модели, такие как генеративно-состязательные сети (GAN), используются для создания реалистичных синтетических данных. Это полезно для обучения моделей безопасности на случайных или недоступных данных [2]. Например, генеративные модели могут быть использованы для создания синтетических сетевых пакетов, которые затем могут быть использованы для обучения систем обнаружения сетевых атак.

Применение этих методов глубокого обучения позволяет создавать более точные и эффективные системы обнаружения угроз, что играет важную роль в обеспечении информационной безопасности организаций.

Заключение

Это исследование глубоко исследовало применение методов машинного обучения в контексте обеспечения информационной безопасности. Мы рассмотрели разнообразные методы, включая классификацию, кластеризацию, регрессию, а также использование обучения на основе данных и обработки больших данных. Эти методы представляют собой мощный инструментарий для анализа данных и выявления потенциальных угроз. Применение методов машинного обучения позволяет автоматизировать процессы анализа данных, что особенно важно в контексте информационной безопасности, где скорость реакции на угрозы играет решающую роль. Методы классификации помогают определять типы угроз, кластеризация выявляет связи между ними, а регрессия прогнозирует вероятность их возникновения [2, 3]. Технологии глубокого обучения, такие как сверточные нейронные сети и рекуррентные нейронные сети, позволяют выявлять сложные закономерности и неявные зависимости в данных, что помогает в обнаружении новых, ранее неизвестных угроз. Перенос обучения и генеративные модели дополняют этот арсенал, позволяя использовать знания из одной области для улучшения результатов в другой [1]. В заключении, применение методов машинного обучения имеет огромный потенциал для улучшения уровня информационной безопасности. Однако, для достижения максимальной эффективности необходимо учитывать особенности данных, выбирать подходящие модели и постоянно совершенствовать методы анализа. Дальнейшие исследования и разработки в этой области будут способствовать созданию более безопасных и защищенных сред.

Литература

[1] Shoniregun, C. A., & Makanjuola, J. K. (Eds.). (2020). "Handbook of Research on Machine Learning Applications in Cyber Security." IGI Global.

[2] Alazab, M., & Mesleh, A. (Eds.). (2021). "Deep Learning Techniques for Cyber Security." Springer.

[3] Abraham, A., & Cherukuri, A. (Eds.). (2020). "Machine Learning for Cyber Physical Systems: Selected Papers from the International Conference ML4CPS 2019." Springer.

[4] Moustafa, N., & Slay, J. (2016). "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, 25(1-3), 18-31.

[5] Marsono, M. N., & Ismail, R. (2020). "Machine Learning in Security Informatics: Intelligent Techniques in Security Informatics." Springer.

ROBOTOTEXNİKANIN MÜXTƏLİF SƏNAYE SAHƏLƏRİNDƏ TƏTBİQİ

Canməmmədova R.R., Yuryeva E.L.,
Məhərrəmova M.R., Mehdiyeva A.R.
Maka, Ekologiya İnstitutu, Təbii Ehtiyatların Kosmik Tədqiqi İnstitutu

Abstact

Müasir dünyada robot sistemləri istehsal proseslərinin tərkib hissəsinə çevrilmişdir. Qabaqcıl texnologiya ilə təchiz edilmiş robotlar müxtəlif sənaye sahələrinin avtomatlaşdırılmasında iştirak edir. Onların arasında sənaye robotları, tibbi robotlar, təhlükəli şəraitdə əməliyyatları yerinə yetirmək üçün robotlar və bir çox başqaları var. Sənayedə robotlar avtomobil istehsalı, elektronika, qida emalı, metallurgiya və digər sahələrdə istifadə edilə bilər. Bu, məhsuldarlığı yüksəltməyə, məhsulun keyfiyyətini yaxşılaşdırmağa və işçilər üçün riskləri azaltmağa imkan verir.