Вопросы безопасности
Правильная ссылка на статью:
Никитин П.В., Горохова Р.И. Анализ современных интеллектуальных методов защиты критической информационной инфраструктуры // Вопросы безопасности. 2024. № 3. DOI: 10.25136/2409-7543.2024.3.69980 EDN: EXGKAV URL: https://nbpublish.com/library_read_article.php?id=69980

Анализ современных интеллектуальных методов защиты критической информационной инфраструктуры

Никитин Петр Владимирович

ORCID: 0000-0001-8866-5610

кандидат педагогических наук

доцент; кафедра искусственного интеллекта; Финансовый университет при Правительстве Российской Федерации

125993, Россия, г. Москва, Ленинградский пр-т, 49

pvnikitin@fa.ru

Горохова Римма Ивановна

кандидат педагогических наук

доцент; кафедра информационных технологий; Финансовый университет при Правительстве Российской Федерации

125167, Россия, г. Москва, Ленинградский пр-т, 49

rigorokhova@fa.ru

Статья из рубрики "Технологии и методология в системах безопасности"

DOI:

10.25136/2409-7543.2024.3.69980

EDN:

EXGKAV

Дата направления статьи в редакцию:

27-02-2024

Дата публикации:

24-09-2024

Аннотация: Критическая информационная инфраструктура (КИИ), в том числе и финансового сектора, играет ключевую роль в обеспечении устойчивого

функционирования экономических систем и финансовой стабильности государств. Однако растущая цифровизация финансовой отрасли и внедрение инновационных технологий открывают новые векторы атак для злоумышленников. Современные кибератаки становятся все более изощренными, а традиционные средства защиты оказываются неэффективными против новых, ранее неизвестных угроз. Возникает острая необходимость в более гибких и интеллектуальных системах обеспечения кибербезопасности. Таким образом, предметом исследования являются современные интеллектуальные методы и технологии защиты критической информационной инфраструктуры (КИИ) от кибератак. Объектом исследования выступаютметоды и средства обеспечения защиты критической информационной инфраструктуры с использованием технологий искусственного интеллекта и машинного обучения.

Методологической основой данного исследования является комплексный анализ научной литературы, посвященной применению интеллектуальных методов и технологий для защиты критической информационной инфраструктуры и проведенный авторами эксперимент по обнаружению финансовых мошенничеств средствами искусственного интеллекта. В ходе проведенного обзора и критического анализа соответствующих научных публикаций были выявлены ключевые проблемы и нерешенные задачи, требующие дальнейших научных изысканий и практических разработок в данной предметной области, которые были доказаны экспериментально. Основные направления научной новизны: 1. Детальное рассмотрение перспективных подходов на основе технологий искусственного интеллекта и машинного обучения для обеспечения эффективной защиты КИИ организаций от современных сложных кибератак. 2. Выявление и анализ ряда ключевых научно-технических проблем, требующих решения для повышения надежности, интерпретируемости и доверия к интеллектуальным системам кибербезопасности, включая вопросы обеспечения робастности к атакам, активного онлайн-обучения, федеративной и дифференциально-приватной обработки данных. 3. Экспериментальная проверка выявленных проблем с использованием средств машинного и глубокого обучения. 4. Определение перспективных направлений дальнейших исследований и разработок в области применения специализированных методов безопасного и доверенного ИИ для защиты критически важной финансовой инфраструктуры. Таким образом, данное исследование вносит значимый вклад в развитие научно-методического аппарата и практических решений по применению интеллектуальных методов для обеспечения кибербезопасности.

Ключевые слова:

Критическая информационная инфраструктура, информационная безопасность, искусственный интеллект, машинное обучение, глубокое обучение, кибербезопасность, кибератаки, нейронные сети, интеллектуальные методы, DDoS-атаки

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета

Введение

Критическая информационная инфраструктура (КИИ) финансового сектора играет ключевую роль в обеспечении устойчивого функционирования экономических систем и финансовой стабильности государств. Сбои или нарушения работы банковских систем, платежных сервисов, бирж и других финансовых организаций могут иметь катастрофические последствия для благосостояния общества. В то же время растущая

цифровизация финансовой отрасли и повсеместное внедрение инновационных технологий открывают новые векторы атак для злоумышленников.

Современные кибератаки становятся все более изощренными, используя методы социальной инженерии, эксплуатацию неизвестных уязвимостей нулевого дня, распределенные ботнеты и инструменты для автоматического сканирования сетей. Традиционные средства защиты, основанные на сигнатурном анализе и правилах, зачастую оказываются неэффективными против новых, ранее неизвестных угроз. Возникает острая необходимость в более гибких и интеллектуальных системах обеспечения кибербезопасности.

Методы искусственного интеллекта, машинного обучения и интеллектуального анализа данных открывают принципиально новые возможности для защиты критической инфраструктуры финансового сектора. Эти технологии позволяют автоматически обнаруживать аномалии и признаки атак в больших объемах разнородных данных, выявлять сложные закономерности, строить прогнозные модели кибератак и динамически адаптироваться к постоянно меняющимся угрозам.

Интеллектуальные методы способны обрабатывать различные типы киберданных - сетевой трафик, логи событий, телеметрию оборудования, данные SIEM-систем, информацию из открытых источников и многое другое. Применяются передовые технологии глубокого обучения, включая сверточные и рекуррентные нейронные сети, трансформеры, генеративные состязательные сети, методы кластеризации и визуального анализа. Это позволяет создавать комплексные решения нового поколения для обнаружения вторжений, интеллектуального мониторинга и активной защиты критических систем.

Актуальность исследований в области применения интеллектуальных методов защиты КИИ финансового сектора обусловлена несколькими ключевыми факторами:

1. Постоянно растущая сложность и динамичность среды кибербезопасности, требующая адаптивных интеллектуальных решений.

2. Необходимость эффективного анализа больших объемов разнородных данных для своевременного обнаружения аномалий.

3. Обеспечение защиты критических систем в условиях роста числа потенциально возможных векторов атак.

4. Возможность прогнозирования и предотвращения кибератак на ранних стадиях с помощью анализа больших данных.

5. Потребность в интерпретируемых и объясняемых решениях в области кибербезопасности финансового сектора.

Таким образом, исследования и разработка интеллектуальных методов защиты критической инфраструктуры финансового сектора является важнейшей междисциплинарной задачей, имеющей стратегическое значение для обеспечения национальной кибербезопасности и экономического благосостояния государства.

Материал и методы исследования

Методологической основой данного исследования является комплексный анализ научной литературы, посвященной применению интеллектуальных методов и технологий для защиты критической информационной инфраструктуры. Рассмотрим, как

представлена данная тема в различных исследованиях как отечественных, так и зарубежных авторов.

Статья Горбатова В.С. с соавторами[1] акцентирует внимание на необходимости многоуровневой защиты сетевого периметра КИИ для эффективного обеспечения кибербезопасности. В ней авторы анализируют вопросы кибербезопасности сетевого периметра объектов критической информационной инфраструктуры (КИИ). Основными угрозами и уязвимости, по мнению авторов, связаны с вредоносным ПО, уязвимостями в ПО, DDoS-атаками и социальной инженерией. Они предлагают комплексный подход, включающий межсетевые экраны, сегментацию сетей, шифрование, обновление ПО и обучение персонала. Так же в статье подчеркивается важность взаимодействия заинтересованных сторон и постоянного совершенствования методов защиты для противодействия киберугрозам.

В работе Зуева В.Н.[2] исследуется применение методов глубокого обучения для обнаружения аномалий в сетевом трафике. Автор отмечает, что традиционные подходы, основанные на правилах и сигнатурах, часто оказываются неэффективными против новых и сложных атак. В качестве альтернативы предлагается использовать сверточную нейронную сеть (CNN) для анализа сетевого трафика в режиме реального времени. Данные трафика представляются в виде двумерных изображений, что позволяет CNN выявлять значимые признаки и закономерности. Проведенные эксперименты на публичных наборах данных показали, что предложенный метод обеспечивает более высокую точность обнаружения аномалий по сравнению с традиционными методами. Несмотря на ряд ограничений, связанных с необходимостью больших объемов данных для обучения и возможностью обхода системы, автор отмечает перспективность применения методов глубокого обучения в задачах кибербезопасности и необходимость дальнейших исследований в этой области.

Вульфин А.М. в своей статье[3] приводит разработку комплексного подхода к оценке рисков безопасности объектов критической информационной инфраструктуры (КИИ) с использованием методов интеллектуального анализа данных. Автор отмечает важность своевременной и точной оценки рисков для обеспечения непрерывности работы КИИ и предотвращения инцидентов. В работе представлен новый подход, основанный на интеграции и анализе данных о событиях безопасности, уязвимостях и инцидентах, накапливаемых в процессе функционирования КИИ. Предложен комплекс моделей и методов, включающих представление и интеграцию разнородных данных, извлечение знаний с помощью машинного обучения и глубокого обучения, оценку рисков с учетом взаимосвязей факторов, а также визуализацию и интерпретацию результатов. Экспериментальная оценка на реальных данных показала, что разработанный подход обеспечивает более точную и полную оценку рисков по сравнению с традиционными методами. Автор отмечает перспективы дальнейших исследований, в том числе интеграцию с системами управления рисками и развитие методов обеспечения безопасности данных, используемых для оценки рисков.

Исследование Ерохина С. Д., Петухова А. Н.[4] проводит разработку нового подхода к управлению безопасностью критических информационных инфраструктур (КИИ), основанного на концепции асимптотического управления. Авторы отмечают сложность и динамичность среды функционирования КИИ, а также необходимость обеспечения их непрерывной работы. Существующие подходы к управлению безопасностью часто недостаточно гибкие и не могут адаптироваться к быстро меняющимся условиям. Предлагаемый подход предполагает постоянную адаптацию и совершенствование

системы безопасности КИИ в соответствии с изменениями среды и новыми угрозами. Он основан на непрерывном мониторинге и анализе данных о событиях безопасности, уязвимостях, инцидентах и других факторах риска с использованием технологий интеллектуального анализа данных. Авторы представляют архитектуру асимптотического управления безопасностью КИИ, включающую компоненты для сбора и интеграции данных, интеллектуального анализа, оценки рисков и принятия решений, а также визуализации и интерпретации результатов. Обсуждаются преимущества предложенного подхода, такие как гибкость и адаптивность системы безопасности, а также его ограничения, связанные с высокими вычислительными требованиями и необходимостью качественных исходных данных.

Статья [5] анализирует проблему обнаружения аномалий в киберфизических системах, особенно в контексте защиты критической инфраструктуры. Автор Vegesna V. V. отмечает, что традиционные методы обнаружения аномалий на основе правил и пороговых значений часто оказываются неэффективными для сложных киберфизических систем с большим объемом разнородных данных. В качестве альтернативы предлагается использовать подходы машинного обучения. В работе проводится анализ различных методов: глубокие нейронные сети для анализа временных рядов, методы изоляционных лесов для выявления выбросов, вероятностные модели и гибридные подходы. Для каждого метода описываются принцип работы, преимущества и ограничения применительно к задаче обнаружения аномалий в киберфизических системах критической инфраструктуры. В качестве примера исследуется система управления электросетями. Проведено экспериментальное сравнение различных методов машинного обучения на реальных данных. Результаты показывают, что наиболее эффективным является гибридный подход, объединяющий несколько техник. Автор делает вывод о перспективности применения методов машинного обучения для защиты критической инфраструктуры и обнаружения аномалий в киберфизических системах. Отмечаются направления дальнейших исследований, включая разработку специализированных методов для конкретных систем и совершенствование методов обработки данных и обеспечения безопасности самих систем машинного обучения.

Статья Selim G. E. I. и др., представленная в журнале «Multimedia Tools and Applications» [6], направлена на исследование применения методов машинного обучения для обнаружения и классификации аномалий в критической промышленной инфраструктуре Интернета вещей (IIoT). Авторы отмечают растущую важность обеспечения безопасности IIoT-систем, управляющих жизненно важными процессами. Традиционные системы обнаружения вторжений часто оказываются неэффективными для IIoT из-за большого объема и разнородности данных, сложной топологии сетей, необходимости работы в режиме реального времени. В качестве решения предлагается использовать алгоритмы машинного обучения для анализа данных IIoT, выявления аномальных событий и их классификации по типам. В работе рассматриваются три основных алгоритма: случайный лес, метод опорных векторов и многослойный персептрон. Описываются математические основы, преимущества и недостатки каждого метода. Для оценки эффективности проведены эксперименты на наборе данных, содержащем реальные показатели работы IIoT-системы управления газоперерабатывающим заводом. Результаты показывают, что алгоритм случайного леса продемонстрировал лучшую точность классификации аномальных событий, достигающую 96%, при сравнительно небольшом времени обучения. Авторы обсуждают возможные пути дальнейшего улучшения производительности и подчеркивают важность защиты критической IIoT-инфраструктуры с помощью методов машинного обучения. Отмечается

необходимость дальнейших исследований с использованием более обширных наборов реальных данных и учетом специфики различных промышленных отраслей.

В статье Pinto A. и др.[7] приведен обзор современных методов и систем обнаружения вторжений, основанных на технологиях машинного обучения, для защиты критической инфраструктуры. Представленный в статье обзор начинается с определения основных угроз безопасности для критических инфраструктурных систем и недостатков традиционных методов обнаружения вторжений. Затем авторы приводят преимущества и ограничения различных типов систем обнаружения вторжений (IDS) - сетевые, узловые и гибридные. Основная часть статьи представляет собой обзор методов машинного обучения, применяемых в современных IDS: методы обучения с учителем, без учителя, гибридные и методы глубокого обучения. Для каждой категории подробно описываются принципы работы алгоритмов, их сильные и слабые стороны в задачах обнаружения вторжений. Отдельное внимание в статье [7] уделяется вопросам предобработки и отбора признаков для работы с данными о сетевом трафике и журналах событий безопасности. Обсуждаются наиболее перспективные методы и открытые проблемы, такие как большие объемы данных, проблема концептного дрейфа и необходимость адаптации к специфике различных критических инфраструктур. Авторы Pinto A. и др. в результате исследования приходят к выводу, что методы глубокого обучения, в особенности сверточные и рекуррентные нейросети, демонстрируют наибольшую точность обнаружения кибератак и аномалий, но для их успешного применения требуются большие наборы качественных данных.

Система для обнаружения и анализа кибер-угроз, направленных на критическую инфраструктуру, с использованием методов машинного обучения представлено в исследовании Aragonés Lozano M., Pérez Llopis I., Esteve Domingo M.[8]. Подход основан на концепции "охоты за угрозами" (threat hunting), предполагающей активный поиск признаков вредоносной активности. Авторы [8] отмечают, что критическая инфраструктура является привлекательной целью для злоумышленников, а традиционные системы обнаружения вторжений часто неэффективны против современных киберугроз. Предлагаемая в статье система состоит из модулей сбора данных, предобработки, машинного обучения и визуализации результатов. Для обнаружения аномалий и классификации угроз авторы используют алгоритмы случайного леса, изолирующего леса и одноклассовой SVM. Система протестирована на реальных данных телекоммуникационной инфраструктуры крупного оператора. Результаты экспериментов показывают высокую эффективность ансамбля машинного обучения в задачах классификации различных типов кибер-угроз, включая обнаружение новых, неизвестных ранее атак. Aragonés Lozano M. и др. обсуждают достоинства предложенного подхода, такие как его способность к обобщению и адаптации к меняющимся угрозам, а также ограничения, связанные с необходимостью постоянного обновления системы по мере появления новых типов атак. Отмечается перспективность методов машинного обучения для эффективного противодействия кибер-угрозам критической инфраструктуры.

Комплексное рассмотрение проблем кибербезопасности критических инфраструктур в контексте современных угроз, в том числе связанных с применением технологий искусственного интеллекта (ИИ) приведено в статье Raval K. J. и др. [9]. Авторы проводят анализ наиболее распространенных кибератак, представляющих угрозу для критических инфраструктурных систем, таких как атаки отказа в обслуживании, вредоносное ПО, атаки на промышленные системы управления, кибершпионаж и др. Особое внимание уделяется новым угрозам, связанным с развитием ИИ. Рассмотренные подходы к

обеспечению кибербезопасности критических инфраструктур с применением методов ИИ и машинного обучения, проанализированные задачи обнаружения вторжений и аномалий, идентификации уязвимостей, реагирования на инциденты, защиты промышленных систем управления и самого ИИ от атак подтверждают необходимость изучения угроз КИИ. Приведенные многочисленные примеры из литературы по использованию ИИ для обеспечения безопасности в различных секторах указывают на актуальность исследования. В статье представлены нерешенные проблемы и будущие направления исследований, включая обеспечение интерпретируемости и объяснимости решений ИИ, развитие методов машинного обучения в условиях противодействующего обучения, разработку стандартов и методологий оценки рисков, а также решение проблем конфиденциальности и этики применения ИИ.

В исследовании [10] Alqudhaibi A. и др. предлагают несколько иной - проактивный подход к прогнозированию кибер-угроз для защиты критической инфраструктуры в контексте Индустрии 4.0. Ключевая идея заключается в анализе мотиваций и поведенческих моделей потенциальных злоумышленников на основе методов машинного обучения. Авторы в своем исследовании подтверждают, что прогнозирование будущих кибератак может быть улучшено за счет понимания побудительных причин и особенностей тех, кто их совершает. Разработанная система состоит из модулей сбора и обогащения данных о кибер-инцидентах, анализа мотиваций атакующих, прогнозирования угроз и оценки рисков. Для прогнозирования используются алгоритмы машинного обучения, такие как случайный лес и нейронные сети, обучаемые на исторических данных об атаках и мотивациях. Результаты экспериментальной оценки показали, что предложенный подход демонстрирует более высокую точность прогнозирования угроз по сравнению с традиционными методами, достигая около 85-90% точности для разных типов атак. В заключении авторы обсуждают ограничения работы, такие как необходимость постоянного обновления данных и возможные неточности в определении реальных мотиваций атакующих. Также отмечается важность соблюдения этических норм при сборе и использовании данных.

Бочков М. В., Васинев Д. А. в своем исследовании[11] предлагают новый подход к моделированию и оценке устойчивости критической информационной инфраструктуры (КИИ) с использованием математического аппарата иерархических гиперсетей и сетей Петри. Устойчивость КИИ имеет чрезвычайно важное значение, поскольку от ее бесперебойной работы зависят многие критически важные системы. Авторы отмечают, что традиционные методы моделирования зачастую не учитывают сложную иерархическую структуру КИИ и взаимосвязи между ее компонентами. Предложенный подход основан на комбинации двух математических моделей: иерархических гиперсетей для представления структуры КИИ и сетей Петри для моделирования динамики ее функционирования, включая возможные сбои, атаки и процессы восстановления. Использование иерархических гиперсетей позволяет учесть сложную топологию КИИ, а сети Петри применяются для описания поведения компонентов в различных ситуациях. На основе разработанной модели авторы предлагают метрики для оценки устойчивости КИИ, такие как вероятность отказа, среднее время безотказной работы и среднее время восстановления. Результаты моделирования устойчивости КИИ на примере энергетической и телекоммуникационной инфраструктуры показывают, что предложенный подход позволяет более точно оценить уязвимости и устойчивость по сравнению с традиционными методами. В заключение Бочков М. В., Васинев Д. А. обсуждают ограничения работы, связанные с необходимостью задания большого количества параметров модели и трудоемкостью вычислений для крупномасштабных инфраструктур, тем не менее отмечается перспектива применения данного подхода для

моделирования других типов сложных систем.

В исследовании Петрова А. Д., Харченко Е. А.[12] рассматривается проблема своевременного обнаружения аномальных состояний серверов, что является важной задачей для обеспечения надежности и безопасности информационных систем. Основная идея предлагаемого авторами метода заключается в применении морфологического анализа для выявления аномалий в мультимодальных временных рядах, характеризующих работу сервера. Предложенная методика включает следующие основные этапы: сбор и предварительная обработка данных мониторинга сервера; преобразование данных в мультимодальный временной ряд векторов признаков; использование морфологического анализа для выявления структурных аномалий; детектирование аномальных состояний на основе коэффициентов структурных аномалий; визуализация и дальнейший анализ выявленных аномалий. Ключевым элементом метода является применение морфологических операций, чувствительных к локальным и глобальным изменениям формы кривых временных рядов. Экспериментальная оценка на реальных данных мониторинга вычислительных узлов показала эффективность предложенного подхода в обнаружении как внезапных, так и длительных аномалий. В заключении обсуждаются преимущества морфологического подхода, такие как отсутствие необходимости в ручном выделении признаков и возможность обнаруживать аномалии различной природы Также рассматриваются ограничения и перспективы дальнейшего развития метода.

Цибизова Т. Ю., Панилов П. А., Кочешков М. А.[13] предлагают подход к мониторингу безопасности критических информационных инфраструктур на основе когнитивного моделирования, которое заключается в следующем:

- построение когнитивной модели, включающей концепты (угрозы, уязвимости, меры защиты) и причинно-следственные связи между ними;

- определение весов связей на основе экспертных оценок или обучения;

- моделирование динамики системы и анализ полученных результатов для выявления критических угроз и уязвимостей;

- мониторинг изменения состояния безопасности во времени и принятие управленческих решений.

Преимущества подхода: возможность учета нечетких и неполных знаний, гибкость модели, наглядность представления. Ограничения: субъективность определения весов, необходимость экспертных знаний. Авторы подчеркивают перспективность использования когнитивного моделирования для обеспечения безопасности критических инфраструктур и предлагают направления дальнейших исследований.

Проблема обеспечения защиты критических информационных инфраструктур от распределенных атак типа «отказ в обслуживании» (DDoS) проанализирована в статье Воеводина В. А. и др.[14]. Авторами предложена методика оценки защищенности автоматизированных систем управления (АСУ) критической инфраструктуры от DDoS-атак на основе имитационного моделирования с использованием метода Монте-Карло.

Воеводин В. А. с соавторами выделяют следующие основные этапы методики:

1. Формирование модели АСУ и модели угроз безопасности в виде сети массового обслуживания.

2. Определение параметров моделей на основе экспертных оценок и статистических данных.

3. Разработка имитационной модели функционирования АСУ под DDoS-атаками.

4. Проведение множества прогонов имитационной модели с различными начальными условиями.

5. Сбор и статистическая обработка результатов моделирования.

6. Оценка защищенности АСУ на основе полученных данных.

7. Выработка рекомендаций по повышению защищенности.

Апробацию методики авторы провели на модельном примере АСУ объекта энергетики, тем не менее отмечаются преимущества использования имитационного моделирования, такие как возможность учета большого числа факторов и получение статистически достоверных оценок и в других областях экономики и промышленности.

Статья Любухина А. С.[15] рассматривает методы анализа и оценки рисков информационной безопасности на основе аппарата нечеткой логики. Автор отмечает, что традиционные методики зачастую опираются на субъективные экспертные оценки, содержащие неопределенности.

Предлагаемая методика включает следующие основные этапы:

1. Определение набора лингвистических переменных для описания рисков.

2. Формирование базы нечетких правил для оценки рисков на основе экспертных знаний.

3. Выбор функций принадлежности для нечетких переменных.

4. Фаззификация входных переменных.

5. Агрегирование нечетких правил с помощью операций нечеткой логики.

6. Дефаззификация выходной переменной (определение конкретного значения риска).

7. Ранжирование и принятие решений по управлению рисками.

В работе приводится пример применения методики для оценки рисков, связанных с компьютерными вирусами. Отмечаются преимущества нечетко-логического подхода, такие как возможность работы с неопределенностью и использование качественных лингвистических оценок. Предлагаемый в исследовании подход подчеркивает перспективность применения методов нечеткой логики для анализа рисков в условиях неопределенности, характерных для области информационной безопасности.

Статья Zhang Y. и др.[16] посвящена проблеме предсказания связей в знаниевых графах, объединяющих требования по обеспечению информационной безопасности и данные об угрозах кибербезопасности. Авторы предлагают новый метод предсказания связей, основанный на распространении информации по ребрам графа (Edge Propagation).

Авторами выделены основные компоненты предлагаемого подхода, среди которых выделяется кодирование узлов и ребер графа в векторные представления, итеративная

передача информации между векторными представлениями узлов через промежуточные ребра, формирование векторных представлений пар узлов, учитывающих их общие связи на разных расстояниях, обучение классификатора (логистической регрессии) на основе этих векторных представлений для предсказания наличия связи.

Эксперименты на реальных и синтетических наборах данных показали, что предложенный метод превосходит другие подходы к предсказанию связей в знаниевых графах. Использование векторных представлений для кодирования семантики, учет структуры связей на разных уровнях, возможность работы с неполными данными и гетерогенными графами, интерпретируемость за счет использования явных пар узлов, предлагаемые исследователями, подтверждают ключевые преимущества метода. В заключение обсуждаются потенциальные направления дальнейших исследований, включая расширение метода для учета динамики и интеграцию с системами принятия решений.

Использование теории графов для визуального анализа и моделирования кибератак также приведено в статье Rabzelj M., Bohak C., Južnič L. Š., Kos A. и Sedlar U.[17]. Авторы предлагают подход, основанный на создании графовой модели, в которой узлы представляют объекты (уязвимости, векторы атак, действия злоумышленников), а ребра отражают возможные переходы между ними в процессе атаки.

В статье предложены ключевые компоненты подхода:

1. Модель графа кибератаки с различными типами узлов и взвешенными ребрами.

2. Алгоритмы для автоматического построения графа на основе входных данных.

3. Методы анализа графа, такие как поиск кратчайших путей, выявление критических узлов, кластеризация.

4. Визуализация графа кибератаки с использованием различных представлений.

5. Интерфейс для аналитиков кибербезопасности.

Эксперименты с реальными данными показали применимость подхода для оценки рисков, планирования защитных мер и анализа последствий кибератак. К основным преимуществам графовой модели Rabzelj M., Bohak C., Južnič L. Š., Kos A. и Sedlar U. относятся:

- наглядная визуализация сложных процессов и взаимосвязей;

- возможность выявления критических компонентов и уязвимых мест;

- поддержка "что-если" анализа и моделирования сценариев;

- интеграция разнородных данных об угрозах из различных источников.

Указанные преимущества позволяют сделать вывод о перспективности дальнейших исследований в данном направлении, включая интеграцию с системами обнаружения вторжений и применение методов машинного обучения.

Статья Ларионова С. Л.[18] посвящена актуальной проблеме противодействия мошенничеству в сфере онлайн финансовых услуг. Автор анализирует основные виды мошеннической деятельности и предлагает комплекс механизмов для их предотвращения, выявления и реагирования. Ключевые элементы подхода включают

строгую идентификацию пользователей, использование скоринговых моделей и алгоритмов машинного обучения, а также средства блокировки мошеннических операций. Особое внимание уделяется повышению общего уровня кибербезопасности и защите конфиденциальных данных клиентов на основе сбалансированного подхода, что позволяет не усложнять процессы предоставления финансовых услуг для добросовестных пользователей.

Berardi D. и др. [19] предлагает подход к вопросам безопасности сетей с жесткими требованиями к временным характеристикам (Time Sensitive Networking, TSN), в частности, проблемам безопасности протокола точной синхронизации времени (Precision Time Protocol, PTP). Исследователи проводят анализ основных угроз безопасности для PTP, включая подмену лидера синхронизации, атаки типа «человек посередине», распределенные атаки и отказ в обслуживании. Предложена таксономия этих угроз на разных уровнях архитектуры. Разработан прототип безопасной реализации PTP с применением криптографических протоколов и механизмов аутентификации на основе сертификатов. Экспериментальная оценка показала приемлемый уровень накладных расходов на обеспечение безопасности. Berardi D. и др. отмечают критическую важность обеспечения безопасности PTP и других компонентов TSN для защиты промышленных систем управления от злонамеренных воздействий. Дальнейшие направления исследований включают интеграцию предложенных механизмов защиты в существующие системы и разработку методов противодействия распределенным атакам на синхронизацию времени.

Авторы Kim T, Pak W. [20] исследуют применение методов глубокого обучения, основанных на трансформерах, для задачи обнаружения сетевых вторжений (Network Intrusion Detection, NID). В статье предложен новый подход, в котором входные сетевые данные преобразуются в изображения и обрабатываются набором параллельных трансформерных сетей. Каждый трансформер специализируется на определенном типе атак или аномалий в сетевом трафике. Выходы от параллельных трансформеров затем объединяются с помощью полносвязных слоев перед получением окончательной классификации. Проведенные эксперименты на наборах данных NSL-KDD и CSE-CIC-IDS2018 демонстрируют превосходство предложенного метода над рядом существующих подходов к NID на основе глубокого обучения. Ключевые преимущества включают высокую эффективность обнаружения различных типов атак, масштабируемость и возможность интерпретации решений модели. Авторы также отмечают ограничения подхода, связанные с высокими вычислительными требованиями и необходимостью предварительной фильтрации данных.

Проведенный анализ показывает необходимость решения вопросов, связанных поиском современных интеллектуальных методов защиты критической информационной инфраструктуры.

Из комплексного анализа научной литературы можно сделать следующие выводы:

1. Применение методов машинного обучения и искусственного интеллекта является ключевым направлением для выявления аномалий, обнаружения кибератак и обеспечения защиты критической инфраструктуры финансового сектора.

2. Широко используются различные технологии глубокого обучения, такие как сверточные нейронные сети, рекуррентные сети, автоэнкодеры и трансформеры. Данные подходы демонстрируют высокую эффективность в задачах обнаружения аномалий и классификации кибератак.

3. Для повышения точности и эффективности систем обнаружения вторжений применяются ансамблевые методы, объединяющие несколько моделей машинного обучения, специализированных на разных типах атак или аспектах безопасности.

4. Предлагаются методы преобразования различных типов данных (сетевой трафик, логи, данные устройств) в форматы, пригодные для анализа с помощью методов компьютерного зрения и обработки изображений.

5. Разрабатываются архитектуры и методы асимптотического управления безопасностью критической инфраструктуры, обеспечивающие адаптивность и самонастройку систем защиты в динамически меняющейся среде угроз.

6. Для комплексной оценки рисков безопасности критических объектов применяются методы нечеткой логики, когнитивного моделирования и иерархических гиперсетей, позволяющие учитывать неопределенность и взаимосвязи между различными факторами риска.

7. Важное внимание уделяется вопросам визуализации и интерпретируемости моделей машинного обучения, используемых для защиты критической инфраструктуры, например, с помощью визуального анализа графов кибератак.

8. Исследуются проактивные подходы к обеспечению кибербезопасности, основанные на анализе мотивов нарушителей и прогнозировании возможных угроз с применением методов искусственного интеллекта.

9. Разрабатываются специализированные решения для обеспечения безопасности финансовых онлайн-услуг, включая механизмы противодействия мошенничеству на основе скоринговых моделей, выявления аномалий и использования биометрических технологий.

10. Анализируются вопросы обеспечения безопасности промышленного Интернета вещей и компонентов критической инфраструктуры в рамках концепции Индустрии 4.0.

11. Исследуются методы защиты протоколов синхронизации времени и других компонентов сетевой инфраструктуры с жесткими требованиями к временным характеристикам с применением криптографических механизмов.

В целом, анализ литературы показал, что интеллектуальные методы на основе машинного обучения и искусственного интеллекта являются ключевыми технологиями для обеспечения эффективной защиты критической инфраструктуры финансового сектора от современных кибератак и угроз безопасности.

Однако, несмотря на значительные достижения в области применения интеллектуальных методов машинного обучения и искусственного интеллекта для защиты критической инфраструктуры финансового сектора, существует ряд проблем и нерешенных задач, которые требуют дальнейших исследований и разработок, среди которых можно выделить основные:

1. Обеспечение интерпретируемости и объяснимости моделей ИИ, используемых для принятия критически важных решений в сфере кибербезопасности. Необходимо развивать методы визуализации, анализа активаций нейронных сетей и извлечения интерпретируемых правил из «черных ящиков» глубокого обучения.

2. Разработка методов оценки доверия и неопределенности к выводам интеллектуальных

систем обнаружения угроз для повышения надежности и принятия взвешенных решений о реагировании.

3. Исследование устойчивости алгоритмов машинного обучения к атакам, направленным на внесение ошибок в процесс принятия решений (adversarial attacks). Необходимы методы обеспечения робастности к таким атакам.

4. Разработка методов активного и онлайн-обучения для своевременной адаптации систем обеспечения безопасности к изменяющимся условиям и новым типам угроз без необходимости полной переобучения.

5. Совершенствование подходов федеративного и дифференциально-приватного обучения для комбинирования разнородных конфиденциальных данных от различных субъектов без раскрытия частной информации.

6. Исследование методов интеграции экспертных знаний в модели машинного обучения путем гибридизации с логическими и продукционными правилами для повышения интерпретируемости и эффективности.

7. Развитие методов автоматического извлечения признаков и представлений на основе неконтролируемого обучения применительно к задачам кибербезопасности.

8. Разработка методов планирования согласованных действий систем активной защиты критической инфраструктуры на основе ситуационного анализа и прогнозных моделей атакующих.

9. Создание масштабируемых платформ для тестирования, верификации и сравнения различных интеллектуальных методов обеспечения безопасности с использованием реалистичных наборов данных и сценариев атак.

10. Исследование вопросов интеграции специализированных методов безопасного и доверенного ИИ с традиционными технологиями защиты информации для построения целостных систем кибербезопасности.

Экспериментальная проверка

Проверим экспериментально данные выводы. На основе методов машинного или глубокого машинного обучения, авторы провели исследование, которое позволяет выявлять мошенничество в финансовой сфере.

В качестве набора данных был использован набор «Credit Card Fraud Detection» (https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud).

Этот набор данных содержит информацию о транзакциях европейских держателей кредитных карт. В нем представлены 284807 транзакций, из которых 492 были мошенническими, что составляет всего 0,172%. Данные включают числовые переменные, полученные с помощью метода главных компонент. Переменные «Время» и «Сумма» не были преобразованы. Переменная «Время» отражает время в секундах между каждой транзакцией и первой, а переменная «Сумма» представляет сумму транзакции. Целевая переменная «Класс» принимает значение 1 в случае мошенничества и 0 в противном случае.

Если решать данную задачу бинарной классификации «в лоб», применив любой из алгоритмов машинного обучения, без анализа данных и за метрику оценки качества применить Accuracy, вычисляемую по формуле (1), то мы получим очень высокий

результат, на уровне 99%, даже если алгоритм не угадает ничего в классе «1».

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \#(1)$$

где TP – true positive, количество верно предсказанных объектов целевого класса;

TN – true negative, количество верно предсказанных объектов нулевого класса;

FP – false positive, количество неверно предсказанных объектов целевого класса;

FN – false negative, количество неверно предсказанных объектов нулевого класса;

Следовательно, действительно необходим анализ данных, выбор методов балансировки классов (или доказательства, что она не нужна), обоснованный выбор метрик оценки обучения, анализ методов машинного и глубокого обучения.

В рамках исследования было протестировано более 20 моделей машинного и глубокого обучения в шести различных сценариях:

1. Без предобработки выбросов и балансировки классов.

2. С предобработкой выбросов без балансировки классов.

3. Без предобработки выбросов с балансировкой через взвешивание классов.

4. С предобработкой выбросов и балансировкой через взвешивание классов.

5. Без предобработки выбросов с применением SMOTE для балансировки.

6. С предобработкой выбросов с применением SMOTE для балансировки.

Основной метрикой оценивания моделей будет ROC-AUC, так как она наиболее точно отражает информацию о True Positive Rate (TPR) и False Positive Rate (FPR) в зависимости от порога принятия решений. При переводе вещественного вывода в бинарный показатель необходимо установить порог, при котором значение 0 переходит в 1; хотя обычно используется порог 0.5, он может быть не оптимальным в случае несбалансированных классов. Для общей оценки моделей, не зависящей от конкретного порога, применяется AUC-ROC, представляющий собой область под кривой, отражающей соотношение TPR (2) и FPR (3):

$$TPR = \frac{TP}{TP + FN}, \#(2)$$

$$FPR = \frac{FP}{FP + TN}, \#(3)$$

TPR отражает полноту, в то время как FPR указывает долю объектов негативного класса, ошибочно классифицированных алгоритмом. В случае идеального классификатора AUC-ROC равен 1 (FPR = 0, TPR = 1). Если классификатор выдает случайные прогнозы, AUC-ROC стремится к 0.5, когда TP и FP равны. Каждая точка на графике соответствует определенному порогу, а площадь под кривой служит индикатором качества алгоритма: чем больше площадь, тем выше качество. Крутизна кривой также важна, поскольку мы стремимся максимизировать TPR и минимизировать FPR, чтобы кривая приближалась к точке (0,1). Пример на рисунке 1 демонстрирует, что текущая модель отстает от идеала, который представлен равнобедренным треугольником, опирающимся на диагональ.

Кроме того, наблюдается, что ошибки на классе 0 могут улучшать определение класса 1. Баланс между TPR и FPR следует определять индивидуально для каждого конкретного случая.

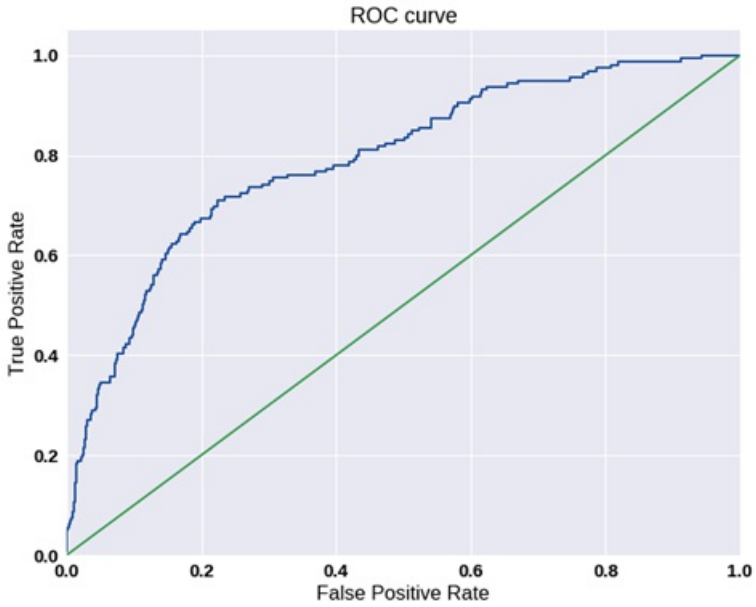

Рисунок 1 - ROC-AUC

Баланс между показателями TPR и FPR является сложной задачей. Снижение порога для идентификации целевого класса «1» может привести к блокировке значительного числа негранычных транзакций, что снизит лояльность клиентов. В то же время минимизация блокировок может повысить риск мошенничества и вызвать недовольство. Поэтому порог принятия решения должен быть определен в зависимости от потребностей банка.

Все модели будут обучаться по единому принципу: методы преобразования будут адаптированы на тренировочной выборке, а на валидационной выборке исключительно применяться. Каждая модель формируется как pipeline из преобразований, с соблюдением сбалансированного деления выборок. Например, если в исходном наборе целевой класс составляет 0,17%, аналогичное соотношение будет соблюдено и для тренировочного, и для валидационного наборов. Первоначальные преобразования включают масштабирование с помощью StandartScaler и RobustScaler.

Для каждой модели будет построен график обучения, отражающий зависимость результатов от размера выборки с указанием доверительного интервала для метрики ROC-AUC. Кривая обучения иллюстрирует влияние объема обучающей выборки на производительность модели и кросс-валидацию, позволяя ответить на два главных вопроса:

1. Как производительность модели изменяется с увеличением объема данных?

2. Насколько модель чувствительна к ошибкам из-за дисперсии по сравнению с ошибками из-за смещения?

Кроме того, для каждой модели будет применяться поиск по сетке для оптимизации гиперпараметров, и обучение будет проводиться только после выбора наилучших значений гиперпараметров.

В таблице 1 отображены результаты протестированных моделей.

Таблица 1 - Метрика ROC-AUC для моделей с разными наборами

Модели	ROC-AUC	Набор без преобразований	Набор с применением SMOTE	Набор со взвешиванием классов
LogisticRegression	с выбросами	0.972423	0.972235	0.973915
LogisticRegression	без выбросов	0.976971	0.975332	0.973418
SVC	с выбросами	0.964920	0.979727	0.977112
SVC	без выбросов	0.962544	0.976860	0.969799
KNeighborsClassifier	с выбросами	0.953505	0.957686	-
KNeighborsClassifier	без выбросов	0.951753	0.957471	-
DecisionTreeClassifier	с выбросами	0.918165	0.90702	0.887544
DecisionTreeClassifier	без выбросов	0.925287	0.924379	0.919717
StackingClassifier	с выбросами	0.943349	0.952068	0.972654
StackingClassifier	без выбросов	0.953258	0.949473	0.977153
BaggingClassifier	с выбросами	0.972654	0.972139	-
BaggingClassifier	без выбросов	0.977153	0.975305	-
AdaboostClassifier	с выбросами	0.918156	0.952031	-
AdaboostClassifier	без выбросов	0.925261	0.954966	-
GradientBoosting	с выбросами	0.927943	0.972651	-
GradientBoosting	без выбросов	0.914239	0.981783	-
NeuralNetwork	с выбросами	0.977698	0.972294	0.972206
NeuralNetwork	без выбросов	0.980729	0.974456	0.976503

Модель KNeighborsClassifier не поддерживает взвешивание классов, при этом наилучшие результаты продемонстрировал метод SMOTE, который оказался практически не зависим от наличия выбросов, хотя оптимальное значение было получено именно с выбросами. С моделями BaggingClassifier, AdaboostClassifier и GradientBoosting также невозможно использовать наборы с взвешиванием классов. Метрика ROC-AUC для BaggingClassifier показала более высокие результаты на наборе данных без каких-либо преобразований. В то же время, метрики ROC-AUC для AdaboostClassifier и GradientBoosting оказались выше при применении SMOTE. Определение лучшей модели происходило на основании метрики ROC-AUC, и вначале была выбрана оптимальная модель среди методов машинного обучения, а затем — лучший подход для нейронной сети. В итоге было отобрано 5 моделей машинного обучения с наилучшими показателями., это:

- (37) GradientBoostingClassifier, обученный на SMOTE наборе, очищенном от выбросов;

- (8) SVC, обученный на SMOTE наборе, не очищенном от выбросов;

- (27) BaggingClassifier обученный на оригинальном наборе, очищенном от выбросов;

- (10) SVC, обученный на оригинальном наборе со взвешиванием классов, не очищенном от выбросов;

- (1) LogisticRegression, обученный на оригинальном наборе данных, очищенном от выбросов.

Для этих моделей был построен график на рисунке 2.

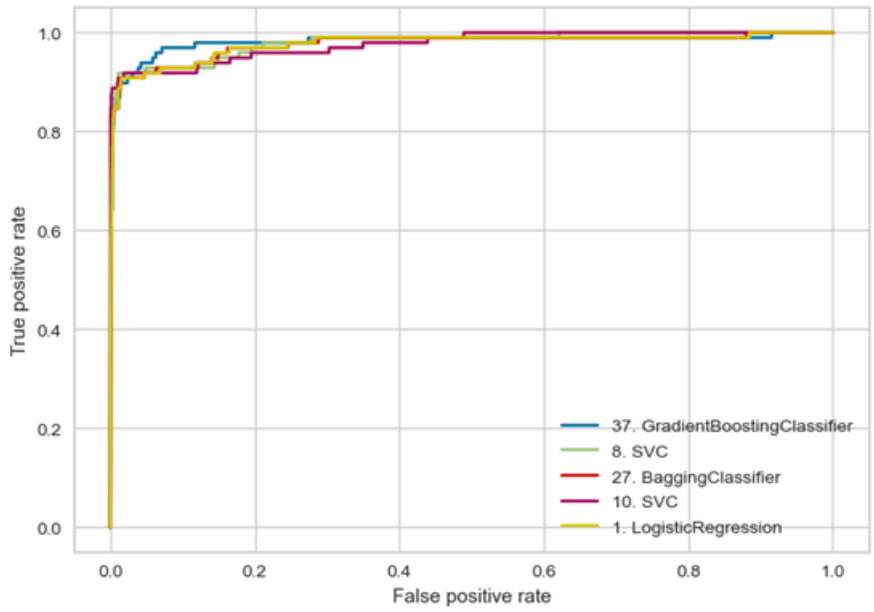

Рисунок 2 – ROC-кривая

В зависимости от того, что важнее:

- выявить все мошеннические операции;

- выявить как можно больше мошеннических транзакций, но минимально ошибочно определить не мошеннические как мошеннические.

Можно выделить двух фаворитов:

- в первом случае это GradientBoostingClassifier 37;

- во втором случае это BaggingClassifier 27 или SVC 8.

Определим, что все-таки большое количество ошибочных выявлений транзакций как мошеннических, больше вредит, чем помогает. Значит выберем все-таки BaggingClassifier 27 как лучшую модель.

Для нейронной сети наилучшие результаты продемонстрировали следующие варианты обработки данных:

- оригинальный набор данных с удалёнными выбросами;

- оригинальный набор данных с сохранёнными выбросами;

- применение взвешивания классов с очищенными выбросами.

Кривая отображена на рисунке 3.

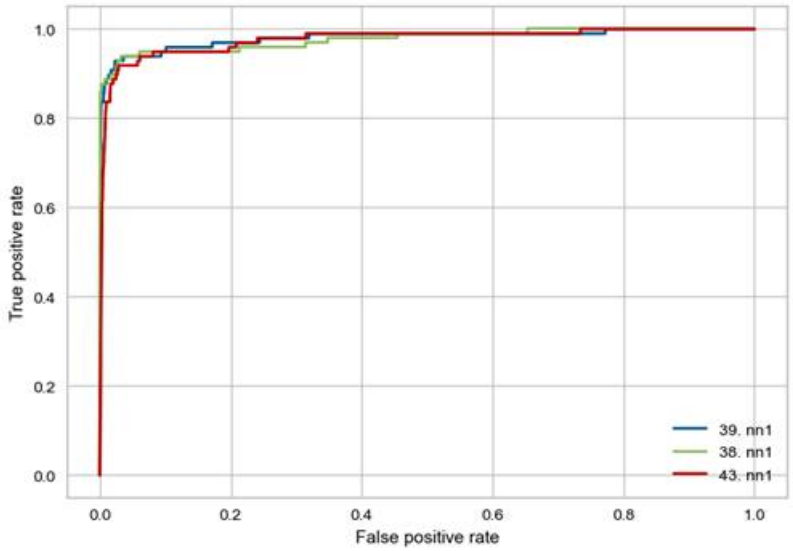

Рисунок 3 – ROC-кривая для nn

Модель 39 показала себя лучше других.

На рисунке 4 представлено финальное сравнение лучших моделей машинного и глубокого обучения.

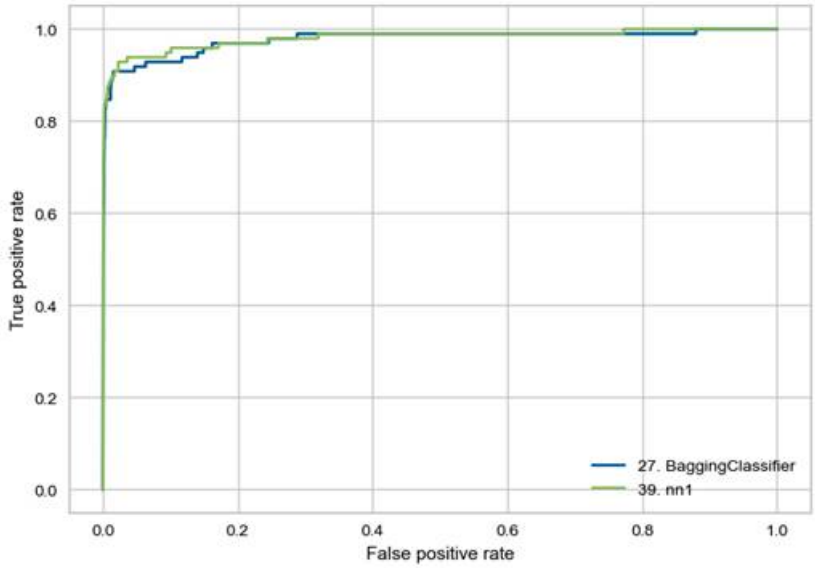

Рисунок 4 – ROC-кривая для nn и BaggingClassifier

Очевидно, что модель нейронной сети демонстрирует лучшие результаты. Поэтому в качестве основной модели будет выбрана нейронная сеть 39, обученная на оригинальном наборе данных, из которого были удалены выбросы.

По результатам экспертизы установлено, что основные научные и практические результаты, полученные в ходе выполнения научно-исследовательской работы, могут быть опубликованы в открытых печатных и электронных научных изданиях.

Результаты и обсуждение

Таким образом, проведенный эксперимент действительно доказывает выделенные

авторами проблемы, описанные выше. Действительно, для принятия критически важных решений в области кибербезопасности необходимы объяснимость моделей ИИ и обеспечение их интерпретируемости. Методы оценки доверия и неопределенности выводов интеллектуальных систем является важнейшей составляющей при разработке алгоритмов защиты КИИ. Необходимо исследовать методы интеграции экспертных знаний в модели машинного обучения путем гибридизации с логическими и продукционными правилами для повышения интерпретируемости и эффективности моделей. Актуальной остается задача создания масштабируемых платформ для тестирования, верификации и сравнения различных интеллектуальных методов обеспечения безопасности с использованием реалистичных наборов данных и сценариев атак. Интеграция специализированных методов безопасного и доверенного искусственного интеллекта с традиционными технологиями защиты информации имеет важное значение для построения целостных систем кибербезопасности.

Проведенное исследование позволяет выделить необходимость применения для решения обозначенных основных проблем интеллектуальных методов защиты критической информационной инфраструктуры финансового сектора:

1. Обеспечение безопасности путем разработки методов и технологий, повышающих уровень безопасности финансовых систем от киберугроз и других рисков. Решение данной проблемы может включать в себя как технические решения (например, системы выявления и предотвращения вторжений, шифрование данных), так и организационные меры (политики информационной безопасности, планы реагирования на инциденты).

2. Создание надежных систем, которые могут эффективно функционировать в условиях различных угроз и загруженности, что необходимо для поддержания стабильности финансового сектора. В первую очередь это интеллектуальные системы защиты, что подразумевает разработку отказоустойчивых архитектур, алгоритмов самовосстановления и адаптации к изменяющимся условиям.

3. Разработка систем искусственного интеллекта понятных и объяснимых для пользователей и специалистов, что касается как работы алгоритмов, так и процесса принятия решений. В системах можно использовать интерпретируемые модели, обеспечивать аудит принятия решений ИИ-системами.

4. Повышение уровня доверия к системам искусственного интеллекта со стороны пользователей и регуляторов. Финансовая отрасль обслуживает жизненно важные интересы граждан, бизнеса и государства. Внедрение ИИ-технологий в этой сфере требует высокого уровня доверия со стороны пользователей и регуляторов, поскольку ошибки или сбои в работе таких систем могут привести к серьезным финансовым и репутационным потерям.

Для повышения доверия к ИИ-системам в финансовой сфере могут быть предприняты следующие меры:

- обеспечение прозрачности и объяснимости работы ИИ-систем. Раскрытие пользователям и регуляторам логики принятия решений ИИ, используемых алгоритмов и источников данных;

- внедрение механизмов аудита и контроля за работой ИИ-систем со стороны независимых экспертов, что позволит подтверждать корректность их функционирования;

- разработка и внедрение строгих этических норм и принципов использования ИИ в

финансовой сфере, это создаст уверенность в том, что системы будут использоваться ответственно и в интересах пользователей;

- активное вовлечение пользователей и регуляторов в процесс разработки и внедрения ИИ-технологий.

Позитивные эффекты от повышения доверия к ИИ-системам в финансах будут направлены на:

- расширение сферы применения ИИ-технологий, повышение их проникновения в критически важные области;

- улучшение качества и надежности финансовых услуг, снижение рисков;

- повышение конкурентоспособности финансовых организаций, использующих ИИ;

- рост инвестиций в развитие ИИ-технологий для финансовой сферы.

Негативные эффекты повлекут за собой:

- дополнительные затраты финансовых организаций на обеспечение прозрачности и контроля за ИИ-системами;

- возможные трудности в достижении полной прозрачности и объяснимости сложных ИИ-систем;

- риск недоверия пользователей к инновациям в финансовой сфере, консервативность.

В целом, повышение доверия к ИИ в финансах за счет комплексных мер является важным и перспективным направлением, способным принести значительные выгоды, но требующим тщательной проработки.

В то же время актуальность подтверждается выявлением путей дальнейших направлений исследования среди которых можно выделить:

- проведение всестороннего анализа угроз и рисков, характерных для критической информационной инфраструктуры финансового сектора, с целью определения приоритетных направлений для разработки методов защиты;

- исследование перспективных технологий искусственного интеллекта (машинное обучение, глубокое обучение, нейронные сети) и их применение для обеспечения безопасности финансовых систем;

- разработка методологий оценки надежности и отказоустойчивости интеллектуальных систем защиты критической информационной инфраструктуры;

- изучение подходов к обеспечению прозрачности и объяснимости работы ИИ-систем в области кибербезопасности финансовых организаций.

Таким образом, несмотря на значительный прогресс, защита критической инфраструктуры финансового сектора с помощью интеллектуальных методов остается активной междисциплинарной областью исследований, требующей решения фундаментальных проблем обеспечения безопасности, надежности, интерпретируемости и доверия к системам искусственного интеллекта, что говорит об актуальности данного фундаментального исследования.

Библиография

1. Горбатов В. С. и др. Кибербезопасность сетевого периметра объекта критической информационной инфраструктуры // Безопасность информационных технологий. 2022. Т. 29. №. 4. С. 12-26.
2. Зуев В. Н. Обнаружение аномалий сетевого трафика методом глубокого обучения // Программные продукты и системы. 2021. Т. 34. №. 1. С. 91-97.
3. Вульфин А. М. Модели и методы комплексной оценки рисков безопасности объектов критической информационной инфраструктуры на основе интеллектуального анализа данных // Системная инженерия и информационные технологии. 2023. Т. 5. №. 4 (13). С. 50-76.
4. Ерохин С. Д., Петухов А. Н. Архитектура асимптотического управления безопасностью критических информационных инфраструктур // DSPA: Вопросы применения цифровой обработки сигналов. 2022. Т. 12. № 1. С. 18-30.
5. Vegesna V. V. Machine Learning Approaches for Anomaly Detection in Cyber-Physical Systems: A Case Study in Critical Infrastructure Protection // International Journal of Machine Learning and Artificial Intelligence. 2024. vol. 5. №. 5. Pp. 1-13.
6. Selim G. E. I. et al. Anomaly events classification and detection system in critical industrial internet of things infrastructure using machine learning algorithms // Multimedia Tools and Applications. 2021. vol. 80. №. 8. Pp. 12619-12640.
7. Pinto A. et al. Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure // Sensors. 2023. vol. 23. №. 5. Pp. 2415.
8. Aragonés Lozano M., Pérez Llopis I., Esteve Domingo M. Threat hunting system for protecting critical infrastructures using a machine learning approach // Mathematics. 2023. vol. 11. №. 16. Pp. 3448.
9. Raval K. J. et al. A survey on safeguarding critical infrastructures: Attacks, AI security, and future directions // International Journal of Critical Infrastructure Protection. 2023. Pp. 100647.
10. Alqudhaibi A. et al. Predicting cybersecurity threats in critical infrastructure for industry 4.0: a proactive approach based on attacker motivations // Sensors. 2023. vol. 23. №. 9. Pp. 4539.
11. Бочков М. В., Васинев Д. А. Моделирование устойчивости критической информационной инфраструктуры на основе иерархических гиперсетей и сетей Петри // Вопросы кибербезопасности. 2024. №. 1. С. 59.
12. Петров А. Д., Харченко Е. А. Морфологический метод обнаружения аномальных состояний сервера // Вестник СибГУТИ. 2023. Т. 18. №. 1. С. 3-15.
13. Цибизова Т. Ю., Панилов П. А., Кочешков М. А. Мониторинг безопасности системы защиты информации критической информационной инфраструктуры на основе когнитивного моделирования // Известия Тульского государственного университета. Технические науки. 2023. №. 6. С. 33-41.
14. Воеводин В. А. и др. Методика оценки защищённости автоматизированной системы управления критической информационной инфраструктуры от DDoS-атак на основе имитационного моделирования методом Монте-Карло // Вестник Дагестанского государственного технического университета. Технические науки. 2023. Т. 50. №. 1. С. 62-74.
15. Любухин А. С. Методы анализа рисков информационной безопасности: нечеткая логика // International Journal of Open Information Technologies. 2023. vol. 11. №. 2. Pp. 66-71.
16. Zhang Y. et al. Edge propagation for link prediction in requirement-cyber threat intelligence knowledge graph // Information Sciences. 2024. vol. 653. Pp. 119770.
17. M. Rabzelj, C. Bohak, L. Š. Južnič, A. Kos and U. Sedlar. Cyberattack Graph Modeling for

Visual Analytics // *IEEE Access.* 2023. vol. 11. Pp. 86910-86944.
18. Ларионова С. Л. Механизмы противодействия мошенничеству в системах онлайн предоставления финансовых услуг // Финансовые рынки и банки. 2023. №. 3. С. 47-52.
19. Berardi D. et al. Time sensitive networking security: issues of precision time protocol and its implementation // Cybersecurity. 2023. vol. 6. №. 1. Pp. 8
20. Kim T, Pak W. Deep Learning-Based Network Intrusion Detection Using Multiple Image Transformers // *Applied Sciences*. 2023. 13(5):2754. https://doi.org/10.3390/app1305275

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Рецензируемая статья посвящена анализу современных интеллектуальных методов защиты критической информационной инфраструктуры.
Методология исследования базируется на комплексном анализе научной литературы, посвященной применению интеллектуальных методов и технологий для защиты критической информационной инфраструктуры.
Актуальность работы обусловлена тем, что растущая цифровизация финансовой отрасли и повсеместное внедрение инновационных технологий открывают новые векторы атак для злоумышленников на критическую информационную инфраструктуру финансового сектора, которая играет ключевую роль в обеспечении устойчивого функционирования экономических систем и финансовой стабильности государств.
Научная новизна рецензируемого исследования, по мнению рецензента, состоит в выявлении проблем и нерешенных задач, которые требуют дальнейших исследований и разработок в сфере интеллектуальных методов защиты критической информационной инфраструктуры.
Структурно в публикации выделены следующие разделы: Введение, Материал и методы исследования, Результаты исследования и их обсуждение, Библиография.
В статье по результатам обобщения литературных источников информации вторами сделан вывод, что интеллектуальные методы на основе машинного обучения и искусственного интеллекта являются ключевыми технологиями для обеспечения эффективной защиты критической инфраструктуры финансового сектора от современных кибератак и угроз безопасности. Среди проблем, требующих дальнейших исследований, авторы в частности отмечают необходимость развития методов визуализации, анализа активаций нейронных сетей и извлечения интерпретируемых правил из «черных ящиков» глубокого обучения, разработки методов оценки доверия и неопределенности к выводам интеллектуальных систем обнаружения угроз и другие.
Библиографический список включает 20 источников – современные публикации на отечественных и зарубежных авторов на русском и английском языках по теме статьи, на которые в тексте имеются адресные ссылки, подтверждающие наличие апелляции к оппонентам.
Из резервов улучшения публикации следует отметить, что в разделе «Материал и методы исследования» при обзоре литературы используются однообразные фразы, например, «статья посвящена…», повторяющаяся 7 раз, а между рассматриваемыми публикациями сложно проследить какую-то логическую связь. По сути, описание источников сводится к изложению её наименования и аннотации – такой стиль изложения трудно назвать удачным. К сожалению, кроме обобщения ранее опубликованных работ, иные методы научного познания в статье не применены, здесь нет ни результатов выборочных обследований, ни каких бы то ни было количественных оценок. Также отсутствуют

конкретные разработки, направленные на решение выявленных проблем – публикация завершается изложением нерешенных вопросов, тогда как обычно в завершении исследований обычно принято приводить сведения о том, какие проблемы удалось решить в результате выполненной работы.
Статья отражает результаты проведенного авторами исследования, соответствует направлению журнала «Вопросы безопасности», но элементы научной новизны и практической значимости в публикации не сформулированы.

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Предмет исследования. Исходя из названия, статья должна быть посвящена анализу современных интеллектуальных методов защиты критической информационной инфраструктуры. Ознакомление с содержанием показало, что оно не противоречит заявленной теме.
Методология исследования базируется как на применении общенаучных методов (анализа, синтеза, индукции и дедукции, графического инструментария), так и специфических (проведение экспериментальной проверки с использованием математического аппарата). Широкий набор применяемых методов формирует позитивное впечатление от ознакомления с рецензируемой научной статьёй.
Актуальность исследования вопросов противодействия кибератакам, не вызывает сомнения, т.к. этот вопрос в настоящее время является одним из важнейших компонентов обеспечения не только информационной безопасности, но и национальной безопасности. Постоянные действия недружественных стран, направленные против отечественной информационной инфраструктуры, говорят о наличии рисков потери информации. Поэтому качественные научные исследования по данной теме будут востребованы у широкой читательской аудитории: данные вопросы находятся и в фокусе внимания органов государственной власти Российской Федерации, и в фокусе внимания конкретных организаций. В тоже время важно учитывать, что отдельные инструменты противодействия угрозам не следует подробно представлять в открытой печати ввиду того, что с ними могут ознакомиться представители недружественных стран и придумать механизмы обхода. Автору рекомендуется в тексте дать оценку возможности опубликования полученных результатов в открытой печати, в т.ч. пояснив невозможность представления отдельных итогов проведённой научной работы.
Научная новизна в представленном на рецензирование материале содержится. В частности, она может быть связана с результатами проведённой экспериментальной проверки. Ценно, что она проводилась на базе иностранных данных, что минимизирует риски возможного применения данных результатов в рамках противодействия российской практики предупреждения угрозк кибербезопасности представителями недружественных стран.
Стиль, структура, содержание. Стиль изложения является научным. Структура статьи выстроена автором, позволяет раскрыть заявленную тему, однако в неё необходимо добавить раздел с рекомендациями решения выявленных проблем и дальнейшими направлениями исследования. Потенциальная читательская аудитория ожидает увидеть в тексте статьи не только перечень существующих проблем, но и конкретные аргументированные мероприятия по их решению. В частности, автор говорит о том, что необходимо «Повышение уровня доверия к системам искусственного интеллекта со стороны пользователей и регуляторов, что является ключевым для их внедрения в

важнейшие сферы, такие как финансы.» Каким образом автор пришёл к выводу о том, что это является ключевым? Как именно нужно повышать доверие? Какие позитивные и негативные эффекты могут быть от реализации данного предложения?
Библиография. Библиографический список состоит из 20 наименований. Ценно, что в нём содержатся как отечественные, так и зарубежные научные публикации. Положительное впечатление формирует и наличие в списке источников публикаций, вышедших в 2024 году.
Апелляция к оппонентам. В тексте автором частично присутствуют элементы апелляции к оппонентам. Также рекомендуется обсудить полученные результаты в виде проблем и рекомендаций по их решению, показав, в чём состоит прирост научного знания?
Выводы, интерес читательской аудитории. С учётом всего вышеизложенного заключаем о том, что статья требует проведения доработки, после чего может быть решён вопрос о целесообразности её опубликования.

Результаты процедуры окончательного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Предметом рецензируемого исследования выступают интеллектуальные методы защиты критической информационной инфраструктуры (далее КИИ) финансового сектора. Авторы справедливо связывают высокую актуальность своего исследования с прогрессирующей цифровизацией социальных процессов вообще, и финансового сектора, в частности, а также с крайне болезненными для благосостояния современных обществ последствиями нарушениями работы финансовых институтов. Дополнительную значимость рецензируемому исследованию придаёт наблюдаемая в последние годы «интеллектуализация» методов защиты КИИ, связанная с расширением применения методов искусственного интеллекта, машинного обучения и интеллектуального анализа данных. В качестве базового метода исследования авторы декларируют «комплексный анализ научной литературы», однако явно не ограничиваются им. Как минимум, результаты анализа литературы проверялись экспериментальным методом. Кроме того, сам эксперимент проводился методами глубокого машинного обучения (о чём ниже говорят сами авторы). Несмотря на некоторую конспективность изложения, обусловленную выбранной авторами методологией исследования, наличие экспериментальной проверки полученных по результатам анализа научной литературы выводов придаёт рецензируемой статье характер научной работы, а не обычного конспекта. Соответственно, можно говорить о научной новизне полученных в процессе исследования результатов. Прежде всего, следует отметить общий вывод о том, что интеллектуальные методы становятся сегодня ключевыми технологиями для обеспечения эффективной защиты от современных кибератак и угроз безопасности. При этом научный интерес представляют также выявленные проблемы и нерешённые задачи в исследуемой области. Наконец, внимания научного сообщества заслуживают результаты проведённого эксперимента, в процессе которого была установлена бОльшая эффективность нейронных сетей в решении задач защиты КИИ, чем традиционные методы. В структурном плане рецензируемая статья также не вызывает существенных нареканий: её логика последовательна и отражает основные аспекты проведённого исследования. В тексте выделены следующие разделы: - «Введение», где ставится научная задача и аргументируется её актуальность; - «Материал и методы исследования», где декларируется (но, к сожалению, не аргументируется) методологическая база исследования, а также проводится анализ научной литературы

по теме; - «Экспериментальная проверка», где проверяются полученные по материалам научной литературы выводы; - «Результаты и обсуждение», где резюмируются итоги проведённого исследования, делаются выводы и намечаются перспективы дальнейших исследований. Стиль статьи научно-аналитический. Текст написан достаточно грамотно, на хорошем русском языке, с корректным использованием научной терминологии. Библиография насчитывает 20 наименований, в том числе источники на иностранных языках, и в должной мере отражает состояние исследований по проблематике статьи. Апелляция к оппонентам имеет место при анализе научной литературы по проблематике статьи. К достоинствам статьи (помимо экспериментальной проверки полученных результатов, что встречается, к сожалению, не так часто) можно отнести использование иллюстративного материала, существенно упрощающего восприятие аргументов авторов.

ОБЩИЙ ВЫВОД: предложенную к рецензированию статью можно квалифицировать в качестве научной работы, отвечающей основным требованиям, предъявляемым к работам подобного рода. Полученные авторами результаты будут интересны для социологов, экономистов, специалистов в области информационной безопасности, а также для студентов перечисленных специальностей. Представленный материал соответствует тематике журнала «Вопросы безопасности». По результатам рецензирования статья рекомендуется к публикации.