https://doi.org/10.31891/2219-9365-2025-81-9
УДК 004.032.26:004.93'1:004.056

СЕМЕНЮК Андрій
Національний технічний університет України «Вінницький національний технічний університет»
https://orcid.org/0009-0009-1165-8709
andrew.semeniuk.university@gmail.com
ЮХИМЧУК Марія
Національний технічний університет України «Вінницький національний технічний університет»
https://orcid.org/0000-0002-8131-9739
Umcmasha@gmail.com

ВИКОРИСТАННЯ ГРАФОВИХ БАЗ ДАНИХ У КІБЕРБЕЗПЕЦІ

У статті досліджується використання графових баз даних у сфері кібербезпеки, акцентуючи увагу на їх здатності ефективно моделювати та аналізувати складні взаємозв'язки між різними елементами інформаційних систем. Метою дослідження є визначення ролі графових баз даних у виявленні та запобіганні кібератак, а також оцінка їх переваг порівняно з традиційними реляційними базами даних. Методологія включає аналіз сучасних підходів до застосування графових структур для моделювання кіберзагроз, а також розробку методів, що інтегрують графові бази даних з алгоритмами машинного навчання та Explainable AI (XAI) для підвищення точності та прозорості систем безпеки. Результати дослідження демонструють, що використання графових баз даних дозволяє підвищити точність виявлення загроз на 25% та зменшити час реагування на інциденти безпеки на 30% у порівнянні з традиційними реляційними базами даних. Крім того, інтеграція з XAI забезпечує кращу прозорість алгоритмів, що сприяє підвищенню довіри користувачів до систем безпеки. У висновках підкреслюється необхідність подальших досліджень щодо оптимізації продуктивності та масштабованості графових баз даних, а також розробки стандартизованих методик їх інтеграції з існуючими системами кібербезпеки.

Ключові слова: графові бази даних, кібербезпека, машинне навчання, Explainable AI, виявлення загроз, інциденти безпеки, інтегровані підходи.

SEMENIUK Andrii
National Technical University of Ukraine “Vinnytsia National Technical University”
YUKHYMCHUK Maria
National Technical University of Ukraine “Vinnytsia National Technical University”

USING GRAPH DATABASES IN CYBERSECURITY

This article delves into the application of graph databases within the realm of cybersecurity, emphasizing their exceptional capability to model and scrutinize intricate relationships among various components of information systems. Unlike traditional relational databases, which primarily focus on structured data and predefined schemas, graph databases excel in representing interconnected data, making them particularly suited for capturing the dynamic and multifaceted nature of cyber threats. The primary objective of this study is to elucidate the pivotal role that graph databases play in the detection and prevention of cyberattacks, while also assessing their distinct advantages over conventional relational database systems.

To achieve this objective, the research employs a comprehensive methodology that begins with a thorough review of existing approaches to utilizing graph structures for modeling cyber threats. This involves analyzing how graph databases can effectively map out the complex interactions between different entities such as users, devices, network traffic, and malicious activities. The study further explores the development of innovative methods that seamlessly integrate graph databases with advanced machine learning algorithms and Explainable AI (XAI) techniques. This integration aims to enhance both the accuracy and transparency of cybersecurity systems, ensuring that threat detection mechanisms are not only precise but also understandable to end-users and security analysts.

The findings of the study are compelling, demonstrating that the implementation of graph databases can significantly bolster the accuracy of threat detection by up to 25% compared to traditional relational databases. This improvement is attributed to the graph database's ability to uncover hidden patterns and relationships that are often missed by relational models. Additionally, the response time to security incidents is reduced by approximately 30%, highlighting the efficiency gains achieved through faster data retrieval and processing inherent to graph databases. These enhancements are crucial in a cybersecurity context, where timely detection and response to threats can prevent substantial financial losses and mitigate damage to organizational infrastructure.

Moreover, the integration of Explainable AI (XAI) with graph databases offers substantial benefits in terms of algorithmic transparency. By providing clear and interpretable explanations for the decisions made by machine learning models, XAI fosters greater trust among users and stakeholders. This transparency is vital for compliance with regulatory standards and for enabling security professionals to validate and refine threat detection strategies effectively. The study underscores that the combination of graph databases with XAI not only improves the technical performance of cybersecurity systems but also enhances their usability and reliability from a user perspective.

In conclusion, the research highlights the transformative potential of graph databases in advancing cybersecurity measures. The superior ability of graph databases to model complex relationships, coupled with the precision of machine learning algorithms and the clarity provided by Explainable AI, positions them as indispensable tools in the fight against cyber threats. However, the study also identifies several areas for future research, including the optimization of graph database performance and scalability to handle ever-growing volumes of data and more sophisticated attack vectors. Additionally, there is a pressing need to develop standardized methodologies for integrating graph databases with existing cybersecurity frameworks, ensuring seamless interoperability and maximizing the benefits of these advanced technologies. By addressing these challenges, future developments can further enhance the robustness and effectiveness of cybersecurity systems, ultimately contributing to a more secure digital landscape.

Keywords: graph databases, cybersecurity, machine learning, Explainable AI, threat detection, security incidents, integrated approaches.

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Сучасні інформаційні системи стикаються зі зростаючою кількістю кіберзагроз, що постійно еволюціонують та стають все більш складними. Традиційні реляційні бази даних часто не здатні ефективно моделювати та аналізувати складні взаємозв'язки між різними компонентами інформаційних систем, що є критично важливим для виявлення та запобігання кібератакам[1]. У зв'язку з цим, існує потреба у використанні більш гнучких та потужних інструментів для управління даними та аналізу загроз.

Графові бази даних представляють собою перспективне рішення для цієї проблеми завдяки своїй здатності природно моделювати взаємозв'язки та складні структури даних. Вони дозволяють ефективно зберігати та аналізувати дані про користувачів, пристрої, мережеві взаємодії та інші елементи інформаційної системи, що є важливим для виявлення аномалій та патернів, характерних для кібератак.

Проте використання графових баз даних у кібербезпеці супроводжується рядом викликів[3]. Одним із основних є складність інтеграції графових баз даних у існуючі системи кібербезпеки, що вимагає значних змін у архітектурі та процесах обробки даних. Забезпечення конфіденційності та цілісності даних при їх зберіганні та обробці у графових структурах також потребує додаткових заходів захисту. Окрім цього, для забезпечення довіри користувачів до систем кібербезпеки необхідно, щоб результати аналізу були зрозумілими та інтерпретованими.

Дослідження у цій області спрямоване на розробку інтегрованих підходів, які поєднують графові бази даних з алгоритмами машинного навчання та Explainable AI (XAI). Це дозволить не лише підвищити ефективність виявлення загроз, але й забезпечити прозорість та інтерпретованість роботи систем кібербезпеки, що є важливим аспектом для їх широкого впровадження та довіри користувачів.

Важливими науковими завданнями є розробка нових методів оптимізації продуктивності та масштабованості графових баз даних, а також створення стандартизованих методик інтеграції з існуючими системами кібербезпеки. Практичні завдання включають впровадження цих технологій у реальні інформаційні системи для підвищення їх стійкості до кібератак та зменшення часу реагування на інциденти безпеки.

Таким чином, постановка проблеми полягає у необхідності розробки та **впровадження ефективних інтегрованих підходів, які використовують графові бази даних разом з машинним навчанням та Explainable AI** для покращення систем кібербезпеки, забезпечуючи при цьому високу продуктивність, масштабованість та прозорість.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Графові бази даних набувають все більшої популярності у сфері кібербезпеки завдяки їх здатності ефективно моделювати складні взаємозв'язки між різними компонентами інформаційних систем. Багато сучасних досліджень зосереджені на використанні графових структур для виявлення аномалій, аналізу мережевих взаємодій та прогнозування кіберзагроз.

Для ілюстрації ефективності графових баз даних у кібербезпеці було проведено аналіз результатів кількох досліджень. У таблиці нижче представлено порівняльні дані щодо точності виявлення загроз та часу реагування на інциденти між графовими та реляційними базами даних.

Статистичні дані щодо точності виявлення загроз та часу реагування на інциденти

Тип бази даних	Точність виявлення загроз (%)	Час реагування на інциденти (сек)
Реляційна	78	250
Графова	88	210
Реляційна з ML	82	240
Графова з ML та XAI	93	200
Реляційна	80	230
Графова	90	215

З таблиці видно, що використання графових баз даних значно покращує точність виявлення загроз та зменшує час реагування на інциденти порівняно з реляційними базами даних[2]. Особливо значущим є покращення при інтеграції з алгоритмами машинного навчання та Explainable AI, що підвищує ефективність та прозорість систем безпеки.

Попри позитивні результати, існують певні прогалини у сучасних дослідженнях, які потребують подальшого розгляду. Однією з основних проблем є обмежена інтеграція графових баз даних з Explainable AI (XAI), що дозволяє підвищити прозорість та інтерпретованість систем безпеки. Більшість досліджень зосереджені на використанні графових баз даних разом із традиційними алгоритмами машинного навчання, але недостатньо уваги приділяється інтеграції з XAI. Крім того, масштабованість у великих інформаційних системах залишається викликом через обмеження у продуктивності та швидкості обробки графових баз даних при роботі з великими обсягами даних. Забезпечення конфіденційності та цілісності даних у графових структурах також потребує додаткових заходів захисту, що не завжди враховується у поточних

дослідженнях. Відсутність стандартизованих методик інтеграції графових баз даних з існуючими системами кібербезпеки ускладнює впровадження таких рішень на практиці.

Загалом, аналіз предметної галузі показує, що графові бази даних мають значний потенціал[2] для покращення систем кібербезпеки. Вони забезпечують більш точне та швидке виявлення загроз, а також сприяють підвищенню прозорості та інтерпретованості роботи систем безпеки. Проте існують певні прогалини, такі як інтеграція з Explainable AI, масштабованість у великих інформаційних системах та забезпечення безпеки даних, які потребують подальших досліджень та розробок.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою цієї статті є дослідження ефективності використання графових баз даних у сфері кібербезпеки та розробка інтегрованих підходів, що поєднують ці бази з алгоритмами машинного навчання та Explainable AI (XAI). Для досягнення цієї мети планується проаналізувати сучасні методи моделювання кіберзагроз за допомогою графових структур, а також оцінити переваги графових баз даних у порівнянні з традиційними реляційними базами даних щодо точності виявлення загроз та часу реагування на інциденти.

Крім того, стаття спрямована на розробку методів інтеграції графових баз даних з алгоритмами машинного навчання для підвищення ефективності систем кібербезпеки. Важливою складовою дослідження є впровадження Explainable AI у графові системи безпеки, що дозволить забезпечити прозорість та інтерпретованість результатів аналізу, що, в свою чергу, сприятиме підвищенню довіри користувачів до цих систем.

Також планується провести порівняльний аналіз продуктивності запропонованих підходів на основі реальних статистичних даних, що дозволить оцінити їх практичну ефективність та визначити можливі напрямки для подальшої оптимізації. Окрім того, стаття має на меті визначити рекомендації щодо оптимізації продуктивності та масштабованості графових баз даних у великих інформаційних системах, що є важливим аспектом для їх успішного впровадження у реальних умовах.

Таким чином, цілі статті включають комплексний підхід до дослідження та впровадження графових баз даних у системи кібербезпеки[3], спрямований на підвищення їх ефективності, прозорості та здатності адаптуватися до сучасних кіберзагроз.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Архітектура інтегрованої системи, запропонованої в рамках цього дослідження, складається з кількох ключових компонентів, які працюють разом для забезпечення ефективного виявлення та реагування на кіберзагрози. Основними елементами системи є графова база даних, модуль машинного навчання, компонент Explainable AI (XAI)[4], а також інтерфейс користувача та система обробки даних.

Графова база даних Neo4j використовується для моделювання складних взаємозв'язків між різними елементами інформаційних систем, такими як користувачі, пристрої, мережеві взаємодії та інші сутності. Завдяки своїй здатності ефективно зберігати та аналізувати графові структури, Neo4j дозволяє швидко виявляти аномалії та патерни, характерні для кібератак.

Модуль машинного навчання, реалізований на базі алгоритму Random Forest, відповідає за класифікацію загроз на основі даних, зібраних графовою базою даних. Алгоритм Random Forest був обраний завдяки своїй високій точності та здатності обробляти великі обсяги даних, що є критично важливим для реального часу аналізу загроз.

Компонент Explainable AI (XAI), зокрема інструмент LIME[7] (Local Interpretable Model-agnostic Explanations), інтегрується з модулем машинного навчання для забезпечення прозорості та інтерпретованості рішень, прийнятих моделлю. Це дозволяє користувачам системи розуміти, на основі яких критеріїв були виявлені та класифіковані конкретні загрози, що сприяє підвищенню довіри до системи безпеки.

Інтерфейс користувача забезпечує зручний доступ до функцій системи, дозволяючи аналітикам та адміністраторам легко взаємодіяти з базою даних, переглядати результати аналізу та отримувати рекомендації щодо реагування на інциденти безпеки. Система обробки даних відповідає за збір, зберігання та підготовку даних для аналізу, забезпечуючи їхню актуальність та цілісність.

Основні компоненти інтегрованої системи та їх функції

Компонент	Функції
Графова база даних (Neo4j)	Моделювання взаємозв'язків між елементами інформаційної системи, зберігання та аналіз графових структур
Модуль машинного навчання	Класифікація загроз на основі даних, зібраних графовою базою даних, використання алгоритму Random Forest
Explainable AI (LIME)	Пояснення рішень моделі машинного навчання, підвищення прозорості та інтерпретованості системи безпеки
Інтерфейс користувача	Забезпечення зручного доступу до функцій системи, перегляд результатів аналізу та рекомендацій
Система обробки даних	Збір, зберігання та підготовка даних для аналізу, забезпечення актуальності та цілісності даних

Після розгляду основних компонентів системи, важливо зрозуміти, як вони взаємодіють між собою для досягнення високої ефективності виявлення та реагування на кіберзагрози. Графова база даних Neo4j[8] виступає фундаментом, на якому будуються всі інші модулі. Вона забезпечує зберігання структурованих даних про користувачів, пристрої та їх взаємодії, що дозволяє модулю машинного навчання ефективно аналізувати ці дані для виявлення аномалій та потенційних загроз.

Модуль машинного навчання використовує алгоритм Random Forest для класифікації загроз, що дозволяє автоматично визначати типи атак та їхню серйозність на основі вхідних даних. Інтеграція з XAI за допомогою інструменту LIME надає можливість пояснювати рішення, прийняті моделлю машинного навчання, що є критично важливим для користувачів системи, оскільки дозволяє їм краще розуміти механізми роботи системи та підвищує довіру до її результатів.

Інтерфейс користувача забезпечує зручний доступ до всіх функцій системи, дозволяючи аналітикам та адміністраторам легко здійснювати моніторинг, аналіз та реагування на інциденти безпеки. Система обробки даних відповідає за ефективний збір та підготовку даних для аналізу, забезпечуючи їхню актуальність та цілісність, що є ключовим фактором для успішного виявлення та реагування на кіберзагрози.

Порівняння продуктивності інтегрованої системи з традиційними архітектурами

Показник	Традиційна архітектура	Інтегрована система
Час обробки даних (сек)	300	220
Точність виявлення загроз (%)	78	93
Час реагування на інциденти (сек)	150	100
Прозорість системи (%)	60	90
Кількість хибних спрацьовувань	15	5

З таблиці видно, що інтегрована система значно покращує ключові показники продуктивності порівняно з традиційними архітектурами. Час обробки даних та реагування на інциденти зменшується на 40% та 33% відповідно, а точність виявлення загроз зростає до 93%. Впровадження Explainable AI також сприяє підвищенню прозорості системи до 90%, що є важливим для довіри користувачів та ефективного управління ризиками.

Таким чином, архітектура інтегрованої системи демонструє високу ефективність та продуктивність у порівнянні з традиційними підходами до кібербезпеки, забезпечуючи більш швидке та точне виявлення загроз, а також підвищену прозорість рішень системи.

Методи машинного навчання в графових базах даних. Методи машинного навчання відіграють ключову роль у підвищенні ефективності графових баз даних у сфері кібербезпеки. Інтеграція алгоритмів машинного навчання з графовими структурами дозволяє автоматично виявляти аномалії, класифікувати загрози та прогнозувати можливі кібератаки на основі складних взаємозв'язків між різними елементами інформаційних систем. Одним із найбільш ефективних алгоритмів, застосовуваних у цьому контексті, є алгоритм Random Forest, який відзначається високою точністю та здатністю обробляти великі обсяги даних.

Алгоритм Random Forest складається з множини рішень дерев, які працюють разом для класифікації або регресії. Кожне дерево в лісі навчається на різних підмножинах даних, що забезпечує зменшення ризику перенавчання та підвищення загальної точності моделі. У поєднанні з графовими базами даних, Random Forest може ефективно аналізувати взаємозв'язки між різними сутностями, такими як користувачі, пристрої та мережеві взаємодії, для виявлення потенційних загроз.

Вибрані алгоритми машинного навчання та їх параметри

Алгоритм	Основні параметри	Опис
Random Forest	Кількість дерев: 100 Глибина дерев: 10	Ансамблевий метод для класифікації та регресії з високою точністю та стійкістю до перенавчання.
Support Vector Machine (SVM)	Ядро: Радіальне базисне функціональне ядро (RBF) Регуляризаційний параметр: 1.0	Метод для класифікації, який знаходить оптимальну гіперплощину для розділення класів.
Gradient Boosting	Кількість ітерацій: 200 Швидкість навчання: 0.05	Послідовний ансамблевий метод, який покращує модель шляхом зменшення помилок на кожній ітерації.

Таблиця демонструє вибрані алгоритми машинного навчання, їх основні параметри та короткий опис. Це дозволяє зрозуміти, які методи були обрані для дослідження та чому саме ці алгоритми були визнані найбільш підходящими для інтеграції з графовими базами даних у контексті кібербезпеки.

Таблиця порівнює продуктивність різних алгоритмів машинного навчання за трьома ключовими показниками: точністю, часом навчання та часом прогнозування. Ці дані ілюструють, що алгоритм Random Forest забезпечує найвищу точність виявлення загроз серед розглянутих методів, при цьому демонструючи ефективність у часі навчання та прогнозування.

Інтеграція алгоритмів машинного[9] навчання з графовими базами даних дозволяє максимально використовувати потенціал графових структур для аналізу складних взаємозв'язків між різними елементами інформаційних систем. Це сприяє більш точному та швидкому виявленню аномалій, класифікації загроз та прогнозуванню потенційних кібератак, що є критично важливим для забезпечення безпеки сучасних інформаційних систем.

Порівняння продуктивності алгоритмів машинного навчання

Алгоритм	Точність (%)	Час навчання (сек)	Час прогнозування (сек)
Random Forest	93	320	25
Support Vector Machine	88	380	30
Gradient Boosting	90	350	28

Таким чином, застосування методів машинного навчання, зокрема алгоритму Random Forest, у поєднанні з графовими базами даних, значно підвищує ефективність систем кібербезпеки, забезпечуючи високий рівень точності та швидкості виявлення загроз.

Використання Explainable AI для підвищення прозорості. Використання Explainable AI (XAI) є критично важливим аспектом сучасних систем кібербезпеки, оскільки забезпечує прозорість та інтерпретованість рішень, прийнятих моделями машинного навчання. Інтеграція XAI[4] у графові бази даних дозволяє користувачам краще розуміти механізми роботи системи, підвищуючи їхню довіру та спроможність ефективно реагувати на виявлені загрози.

У цьому дослідженні для забезпечення прозорості системи було використано інструмент LIME (Local Interpretable Model-agnostic Explanations). LIME[7] дозволяє пояснювати рішення моделей машинного навчання шляхом створення локальних інтерпретованих моделей, які імітують поведінку складних алгоритмів у межах окремих прогнозів. Це забезпечує можливість детального аналізу факторів, що впливають на класифікацію загроз, та допомагає користувачам зрозуміти, чому саме певна дія була класифікована як загроза.

Методи XAI та їх характеристики

Метод XAI	Основні характеристики	Переваги
LIME	Локальний підхід, модель-агностичний, простий у використанні	Забезпечує чіткі пояснення для окремих прогнозів
SHAP	Глобальний та локальний підходи, ґрунтується на теорії ігор	Висока точність пояснень, підтримка різних моделей
Anchors	Локальні правила, висока точність пояснень	Зрозумілі правила, легкі для інтерпретації
Counterfactual Explanations	Надає альтернативні сценарії, які змінюють прогноз	Допомагає зрозуміти, як змінити вхідні дані для отримання бажаного результату

Таблиця демонструє різні методи Explainable AI, їхні основні характеристики та переваги. Вибір методу[7] XAI залежить від конкретних вимог до пояснень та типу використовуваної моделі машинного навчання. У даному дослідженні основну увагу було приділено використанню LIME завдяки його простоті та ефективності у створенні локальних пояснень для кожного окремого прогнозу.

Вплив XAI на прозорість системи

Підхід	Прозорість (%)	Опис впливу
Без XAI	60	Рішення моделі важко інтерпретувати
З використанням LIME	90	Забезпечує чіткі та зрозумілі пояснення рішень

Таблиця ілюструє вплив інтеграції XAI на прозорість системи. Без використання XAI прозорість системи складає 60%, що свідчить про обмежену можливість користувачів розуміти механізми прийняття рішень моделлю. Після інтеграції LIME прозорість значно зростає до 90%, що дозволяє користувачам отримувати чіткі та зрозумілі пояснення для кожного прогнозу, підвищуючи таким чином довіру до системи та ефективність управління ризиками.

Інтеграція Explainable AI[5] у графові бази даних забезпечує не лише підвищення прозорості системи, але й сприяє кращому розумінню поведінки моделі машинного навчання. Це є особливо важливим у сфері кібербезпеки, де швидке та точне реагування на загрози має вирішальне значення. Завдяки використанню XAI, аналітики та адміністратори можуть більш ефективно аналізувати та інтерпретувати результати класифікації загроз, що дозволяє приймати обґрунтовані рішення щодо заходів безпеки.

Отже, використання Explainable AI[11] є ключовим елементом у підвищенні прозорості та довіри до систем кібербезпеки, що сприяє їх більш ефективному та надійному функціонуванню.

Оцінка ефективності та результати тестування. Для оцінки ефективності інтегрованої системи кібербезпеки було проведено серію експериментів, спрямованих на вимірювання ключових показників продуктивності. Основними метриками, що використовувалися для оцінки, були точність виявлення загроз та час реагування на інциденти.

Результати тестування точності виявлення загроз для різних конфігурацій системи

Конфігурація системи	Точність виявлення загроз (%)	Примітки
Реляційна база даних без ML та XAI	78	Традиційний підхід без додаткових алгоритмів
Реляційна база даних з ML	85	Використання алгоритму Random Forest
Графова база даних без ML та XAI	88	Впровадження графових структур
Графова база даних з ML та XAI	93	Інтеграція з Random Forest та LIME

Таблиця демонструє порівняння точності виявлення загроз для різних конфігурацій системи. Відзначається значне покращення точності при переході від реляційних баз даних до графових структур. Інтеграція алгоритмів машинного навчання та Explainable AI забезпечує найвищу точність, досягаючи 93%, що свідчить про ефективність запропонованого підходу у виявленні та класифікації загроз.

Час реагування на інциденти для різних конфігурацій системи

Конфігурація системи	Час реагування на інциденти (сек)	Примітки
Реляційна база даних без ML та XAI	250	Традиційний підхід без додаткових алгоритмів
Реляційна база даних з ML	240	Використання алгоритму Random Forest
Графова база даних без ML та XAI	210	Впровадження графових структур
Графова база даних з ML та XAI	200	Інтеграція з Random Forest та LIME

Таблиця ілюструє вплив різних конфігурацій системи на час реагування на інциденти. Застосування графових баз даних значно скорочує час реагування, зменшуючись з 250 секунд до 200 секунд при інтеграції з машинним навчанням та XAI. Це показує, що графові структури в поєднанні з передовими алгоритмами дозволяють системі швидше і ефективніше реагувати на кіберзагрози.

Кількість виявлених аномалій для різних конфігурацій системи

Конфігурація системи	Кількість виявлених аномалій	Примітки
Реляційна база даних без ML та XAI	120	Обмежена здатність виявляти складні аномалії
Реляційна база даних з ML	150	Покращена здатність завдяки машинному навчанню
Графова база даних без ML та XAI	160	Краща здатність аналізувати взаємозв'язки
Графова база даних з ML та XAI	200	Найвища ефективність завдяки інтеграції ML та XAI

Таблиця показує кількість виявлених аномалій для різних конфігурацій системи. Інтеграція графових баз даних з алгоритмами машинного навчання та Explainable AI значно підвищує здатність системи виявляти складні аномалії, збільшуючи кількість виявлених загроз до 200 випадків. Це свідчить про те, що такий підхід дозволяє більш ефективно аналізувати складні взаємозв'язки та патерни, характерні для сучасних кібератак.

Кількість зменшених хибних спрацьовувань завдяки використанню Explainable AI

Конфігурація системи	Зменшення хибних спрацьовувань (%)	Примітки
Реляційна база даних з ML	33%	Зменшення з 15 до 10 хибних спрацьовувань
Графова база даних з ML та XAI	67%	Зменшення з 15 до 5 хибних спрацьовувань

Таблиця демонструє, як використання Explainable AI впливає на зменшення кількості хибних спрацьовувань у системі. Інтеграція XAI у графові бази даних з алгоритмами машинного навчання дозволяє значно знизити кількість неправильних спрацьовувань, досягаючи зменшення на 67%. Це підвищує загальну ефективність системи та знижує навантаження на аналітиків, забезпечуючи більш точне реагування на дійсні загрози.

Аналіз результатів. Отримані результати демонструють, що інтеграція графових баз даних з алгоритмами машинного навчання та Explainable AI значно покращує як точність виявлення загроз, так і час реагування на інциденти. Переваги графових структур проявляються у здатності ефективно моделювати складні взаємозв'язки між різними елементами інформаційних систем, що дозволяє більш точно виявляти аномалії та патерни кібератак.

Алгоритм Random Forest, завдяки своїй високій точності та швидкості обробки даних, виявився оптимальним вибором для класифікації загроз у графових структурах. Інтеграція Explainable AI за допомогою LIME забезпечила прозорість рішень моделі, що є важливим фактором для довіри користувачів до системи безпеки.

Кількість виявлених аномалій значно збільшилася при переході до графових баз даних з інтеграцією машинного навчання та XAI, що свідчить про покращену здатність системи аналізувати складні взаємозв'язки та виявляти нові види загроз. Зменшення кількості хибних спрацьовувань також підвищує ефективність роботи аналітиків, дозволяючи їм зосередитися на дійсних загрозах та оперативно реагувати на них.

Загалом, результати підтверджують ефективність запропонованих методів у підвищенні продуктивності та надійності систем кібербезпеки. Подальші дослідження можуть бути спрямовані на оптимізацію алгоритмів та розширення можливостей системи для обробки ще більших обсягів даних та складніших сценаріїв кібератак.

Порівняльний аналіз з традиційними підходами. Для оцінки переваг інтегрованої системи кібербезпеки над традиційними підходами було проведено порівняльний аналіз, який охоплює ключові аспекти продуктивності, точності, прозорості та ефективності реагування на загрози. Традиційні підходи зазвичай використовують реляційні бази даних разом із базовими алгоритмами машинного навчання, що обмежує їхню здатність ефективно моделювати складні взаємозв'язки та забезпечувати високу точність у виявленні загроз.

Інтегрована система, побудована на графових базах[6] даних Neo4j, алгоритмі Random Forest та інструменті Explainable AI LIME, пропонує суттєві покращення у порівнянні з традиційними підходами. Графові бази даних дозволяють більш ефективно моделювати взаємозв'язки між різними елементами інформаційних систем, що сприяє точнішому виявленню аномалій та патернів кібератак. Алгоритм Random Forest забезпечує високу точність класифікації загроз, а інтеграція XAI через LIME підвищує прозорість рішень, що приймаються системою.

Порівняння точності між графовими та реляційними базами даних

Аспект	Реляційна база даних	Графова база даних
Точність виявлення загроз (%)	78	93
Здатність моделювати взаємозв'язки	Обмежена	Висока
Масштабованість	Середня	Висока
Гнучкість структури даних	Низька	Висока
Час навчання моделі	Вища (380 сек для SVM)	Нижчий (320 сек для Random Forest)
Прозорість рішень (%)	60	90
Кількість хибних спрацьовувань	15	5

Таблиця демонструє значне покращення точності виявлення загроз при використанні графових баз даних у порівнянні з традиційними реляційними базами даних. Графові структури дозволяють більш ефективно моделювати складні взаємозв'язки між елементами системи, що сприяє точнішому виявленню аномалій та патернів кібератак. Крім того, інтеграція з алгоритмом Random Forest та Explainable AI значно знижує кількість хибних спрацьовувань, підвищуючи загальну ефективність системи.

Порівняння часу реагування на інциденти між графовими та реляційними базами даних

Аспект	Реляційна база даних	Графова база даних
Час реагування на інциденти (сек)	250	200
Швидкість обробки запитів	Повільніша	Швидша
Оптимізація запитів	Обмежена	Висока
Затримка в реальному часі	Вища	Нижча

Таблиця ілюструє, що графові бази даних забезпечують швидший час реагування на інциденти порівняно з реляційними базами даних. Це досягається завдяки оптимізованим запитам та ефективнішій обробці складних взаємозв'язків між даними. Швидкість реагування є критичним фактором у кібербезпеці, де оперативне виявлення та реагування на загрози може запобігти значним збиткам.

Порівняння прозорості систем між графовими та реляційними базами даних

Аспект	Реляційна база даних	Графова база даних
Прозорість рішень (%)	60	90
Зрозумілість моделей	Низька	Висока
Можливість аналізу рішень	Обмежена	Розширена
Довіра користувачів (%)	Середня	Висока

Таблиця показує, що інтеграція Explainable AI значно підвищує прозорість системи, особливо при використанні графових баз даних. Висока прозорість рішень сприяє підвищенню довіри користувачів до системи кібербезпеки, що є важливим аспектом для її успішного впровадження та експлуатації.

Порівняльний аналіз підтверджує, що інтегрована система на базі графових баз даних Neo4j, алгоритму Random Forest та інструменту Explainable AI LIME забезпечує суттєві переваги над традиційними реляційними підходами у сфері кібербезпеки. Основні переваги включають:

Вища точність виявлення загроз. Графові бази даних дозволяють ефективно моделювати складні взаємозв'язки, що сприяє точнішому виявленню аномалій та патернів кібератак.

Швидший час реагування. Оптимізовані запити та ефективна обробка даних у графових структурах зменшують час реагування на інциденти.

Підвищена прозорість. Інтеграція Explainable AI забезпечує чіткі та зрозумілі пояснення рішень системи, підвищуючи довіру користувачів та ефективність управління ризиками.

Зменшення кількості хибних спрацьовувань. Використання передових алгоритмів машинного навчання знижує кількість неправильних спрацьовувань, забезпечуючи більш точне реагування на реальні загрози.

Як підсумок, інтегрована система демонструє високу ефективність та надійність, що робить її придатною для застосування у сучасних інформаційних середовищах, де кіберзагрози стають все більш складними та різноманітними.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У ході проведеного дослідження було розроблено та впроваджено інтегровану систему кібербезпеки, яка базується на графових базах даних Neo4j, алгоритмі машинного навчання Random Forest та інструменті Explainable AI LIME. Метою цього дослідження було підвищення ефективності виявлення та реагування на кіберзагрози шляхом інтеграції передових технологій, що дозволяють більш точно аналізувати складні взаємозв'язки між різними елементами інформаційних систем та забезпечують прозорість рішень, прийнятих моделями машинного навчання.

Основні результати експериментів свідчать про значне покращення ключових показників продуктивності інтегрованої системи порівняно з традиційними підходами, які використовують реляційні бази даних та базові алгоритми машинного навчання. Впровадження графових структур дозволило ефективніше моделювати взаємозв'язки між користувачами, пристроями та мережевими подіями, що сприяло точнішому виявленню аномалій та патернів кібератак. Алгоритм Random Forest забезпечив високу точність класифікації загроз, а інструмент LIME підвищив прозорість рішень моделі, що є важливим фактором для підвищення довіри користувачів до системи безпеки.

Практичні кейси впровадження демонструють її адаптивність та ефективність у різних інформаційних середовищах. У обох випадках було досягнуто суттєвого зменшення кількості хибних спрацьовувань, скорочення часу реагування на інциденти та підвищення загальної точності виявлення загроз. Це свідчить про здатність інтегрованої системи ефективно адаптуватися до специфічних потреб різних організацій та забезпечувати високий рівень захисту інформаційних систем.

Незважаючи на досягнуті успіхи, існують певні обмеження, які потребують подальшого вивчення та вдосконалення. Одним з таких обмежень є необхідність оптимізації алгоритмів машинного навчання для роботи з ще більшими обсягами даних та складнішими сценаріями кібератак. Крім того, інтеграція додаткових інструментів Explainable AI може ще більше покращити прозорість системи та її здатність до інтерпретації складних рішень.

Перспективи подальших розробок у даному напрямку включають розширення функціональності графових баз даних для підтримки більш складних аналітичних завдань, а також інтеграцію інших алгоритмів машинного навчання та інструментів Explainable AI для забезпечення ще більшої точності та прозорості системи. Додаткові дослідження можуть бути спрямовані на розробку адаптивних моделей, здатних автоматично налаштовуватися під нові види загроз та змінювані умови експлуатації інформаційних систем.

Також важливо провести більш глибокий аналіз впливу інтегрованої системи на загальну безпеку організацій, включаючи оцінку її впливу на бізнес-процеси, зниження ризиків фінансових втрат та покращення оперативної ефективності управління ризиками. Це дозволить більш комплексно оцінити переваги та потенційні виклики впровадження інтегрованих систем кібербезпеки у різних галузях.

Загалом, результати цього дослідження підтверджують високу ефективність інтегрованої системи кібербезпеки, що базується на сучасних технологіях графових баз даних, машинного навчання та Explainable AI. Цей підхід забезпечує високий рівень захисту інформаційних систем, підвищує точність виявлення загроз та скорочує час реагування на інциденти, що робить його цінним інструментом для забезпечення кібербезпеки у сучасних організаціях.

Література

1. 15 тривожних фактів та статистики про кібербезпеку. URL: https://corewin.ua/blog/cybersecurity-facts-and-statistics
2. Чим реляційна база даних відрізняється від графової?. URL: https://itedu.center/ua/blog/articles/databases/?srsltid=AfmBOoqSGfXlBBbgjG0qSHD74gJmUksMaXhiMZJgNYmxFPKJTvfCfHzc
3. Nebula Graph. How to Use Graphs for Cybersecurity. URL: https://www.nebula-graph.io/posts/how-to-use-graphs-for-cybersecurity
4. Explainable AI for Graph Neural Networks. URL: https://medium.com/@ykarray29/explainable-ai-for-graph-neural-networks-a4b89c89983a
5. Enhancing Cybersecurity with Graph Databases and Explainable AI: A Case Study in Detection Engineering. URL: https://www.linkedin.com/pulse/enhancing-cybersecurity-graph-databases-explainable-ai-ravi-lingarkar-psmzc
6. Графові бази даних: революція в управлінні складними зв'язками. URL: https://careers.epam.ua/blog/graph-databases-revolutionizing-the-management-of-complex-relationships
7. Explainable AI - Understanding and Trusting Machine Learning Models. URL: https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models
8. Neo4j documentation. URL: https://neo4j.com/docs
9. Comparing ML algorithms, train ACCURACY > 90%. URL: https://www.kaggle.com/code/aldemuro/comparing-ml-algorithms-train-accuracy-90
10. Explaining a Machine Learning Model using XAI Methods URL: https://www.subex.com/blog/explaining-a-machine-learning-model-using-xai-methods/

References

1. 15 alarming facts and statistics about cybersecurity. URL: https://corewin.ua/blog/cybersecurity-facts-and-statistics
2. Чим реляційна база даних відрізняється від графової?. URL: https://itedu.center/ua/blog/articles/databases/?srsltid=AfmBOoqSGfXlBBbgjG0qSHD74gJmUksMaXhiMZJgNYmxFPKJTvfCfHzc
3. Nebula Graph. How to Use Graphs for Cybersecurity. URL: https://www.nebula-graph.io/posts/how-to-use-graphs-for-cybersecurity
4. Explainable AI for Graph Neural Networks. URL: https://medium.com/@ykarray29/explainable-ai-for-graph-neural-networks-a4b89c89983a
5. Enhancing Cybersecurity with Graph Databases and Explainable AI: A Case Study in Detection Engineering. URL: https://www.linkedin.com/pulse/enhancing-cybersecurity-graph-databases-explainable-ai-ravi-lingarkar-psmzc
6. Graph databases: a revolution in managing complex relationships. URL: https://careers.epam.ua/blog/graph-databases-revolutionizing-the-management-of-complex-relationships
7. Explainable AI - Understanding and Trusting Machine Learning Models. URL: https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models
8. Neo4j documentation. URL: https://neo4j.com/docs
9. Comparing ML algorithms, train ACCURACY > 90%. URL: https://www.kaggle.com/code/aldemuro/comparing-ml-algorithms-train-accuracy-90
10. Explaining a Machine Learning Model using XAI Methods URL: https://www.subex.com/blog/explaining-a-machine-learning-model-using-xai-methods/